Fourth Edition

LINEAR ALGEBRA
AND ITS APPLICATIONS

Gilbert Strang

5.5 Complex Matrices

It is no longer possible to work only with real vectors and real matrices In the first half of this book, when the basic problem was $Ax - b$, the solution was real when A and b were real. Complex numbers could have been permitted. but would have contributed nothing new. Now we cannot avoid them. A real matrix has real coefficients in $\det(A - \lambda I)$, but the eigenvalues (as in rotations) may be complex.

We now introduce the space $\mathbf{C}^n$ of vectors with *n complex* components. Addition and matrix multiplication follow the same rules as before. ***Length is computed differently***. The old way, the vector in $\mathbf{C}^2$ with components $(1, i)$ would have zero length: $1^2 + i^2 = 0$, not good. The correct length squared is $1^2 + |i|^2 = 2$.

This change to $\|x\|^2 = |x_1|^2 + \cdots + |x_n|^2$ forces a whole series of other changes. The inner product, the transpose, the definitions of symmetric and orthogonal matrices, all need to be modified for complex numbers. The new definitions coincide with the old when the vectors and matrices are real. We have listed these changes in a table at the end of the section. and we explain them as we go.

That table virtually amounts to a dictionary for translating real into complex. We hope it will be useful to the reader. We particularly want to find out about ***symmetric matrices*** and ***Hermitian matrices***: *Where are their eigenvalues, and what is special about their eigenvectors*? For practical purposes, those are the most important questions in the theory of eigenvalues. We call attention in advance to the answers:

1. ***Every symmetric matrix (and Hermitian matrix) has real eigenvalues.***
2. ***Its eigenvectors can be chosen to be orthonormal.***

Strangely, to prove that the eigenvalues are real we begin with the opposite possibility—and that takes us to complex numbers, complex vectors, and complex matrices.

Complex Numbers and Their Conjugates

Probably the reader has already met complex numbers; a review is easy to give. The important ideas are the *complex conjugate* $\bar{x}$ and the *absolute value* $|x|$. Everyone knows that whatever i is, it satisfies the equation $i^2 = -1$. It is a pure imaginary number, and so are its multiples ib; b is real. The sum $a + ib$ is a complex number, and it is plotted in a natural way on the complex plane (Figure 5.4).

The real numbers a and the imaginary numbers ib are special cases of complex numbers; they lie on the axes. Two complex numbers are easy to add:

$$\textbf{Complex addition} \qquad (a + ib) + (c + id) = (a + c) + i(b + d).$$

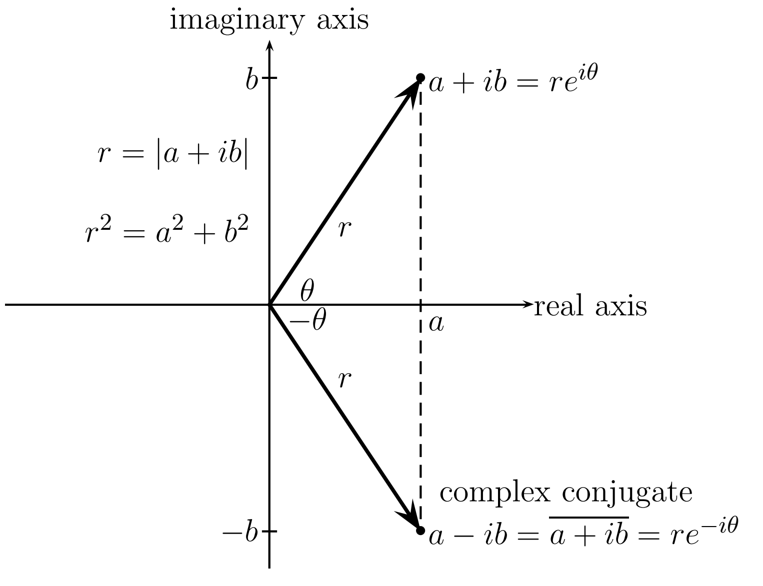

Figure 5.4: The complex plane, with $a+ib = re^{i\theta}$ and its conjugate $a-ib = re^{-i\theta}$.

Multiplying $a+ib$ times $c+id$ uses the rule that $i^2 = -1$:

$$\textbf{Multiplication} \qquad \begin{aligned}(a+ib)(c+id) &= ac+ibc+iad+i^2bd \\ &= (ac-bd)+i(bc+ad).\end{aligned}$$

The ***complex conjugate*** of $a+ib$ is the number $a-ib$. *The sign of the imaginary part is reversed.* It is the mirror image across the real axis; any real number is its own conjugate, since $b=0$. The conjugate is denoted by a bar or a star: $(a+ib)^* = \overline{a+ib} = a-ib$. It has three important properties:

1. The conjugate of a product equals the product of the conjugates:

$$\overline{(a+ib)(c+id)} = (ac-bd)-i(bc+ad) = \overline{(a+ib)}\,\overline{(c+id)}. \tag{1}$$

2. The conjugate of a sum equals the sum of the conjugates:

$$\overline{(a+c)+i(b+d)} = (a+c)-i(b+d) = \overline{(a+ib)} + \overline{(c+id)}.$$

3. Multiplying any $a+ib$ by its conjugate $a-ib$ produces a real number a^2+b^2:

$$\textbf{Absolute value} \qquad (a+ib)(a-ib) = a^2+b^2 = r^2. \tag{2}$$

This distance r is the ***absolute value*** $|a+ib| = \sqrt{a^2+b^2}$.

Finally, trigonometry connects the sides a and b to the hypotenuse r by $a = r\cos\theta$ and $b = r\sin\theta$. Combining these two equations moves us into polar coordinates:

$$\textbf{Polar form} \qquad a+ib = r(\cos\theta + i\sin\theta) = re^{i\theta}. \tag{3}$$

The most important special case is when $r = 1$. Then $a+ib$ is $e^{i\theta} = \cos\theta + i\sin\theta$. It falls on the ***unit circle*** in the complex plane. As θ varies from 0 to 2π, this number $e^{i\theta}$ circles around zero at the constant radial distance $|e^{i\theta}| = \sqrt{\cos^2\theta + \sin^2\theta} = 1$.

Example 1. $x = 3+4i$ times its conjugate $\bar{x} = 3-4i$ is the absolute value squared:

$$x\bar{x} = (3+4i)(3-4i) = 25 = |x|^2 \quad \text{so} \quad r = |x| = 5.$$

To divide by $3+4i$, multiply numerator and denominator by its conjugate $3-4i$:

$$\frac{2+i}{3+4i} = \frac{2+i}{3+4i}\frac{3-4i}{3-4i} = \frac{10-5i}{25}.$$

In polar coordinates, multiplication and division are easy:

$re^{i\theta}$ times $Re^{i\alpha}$ has absolute value rR and angle $\theta+\alpha$.

$re^{i\theta}$ divided by $Re^{i\alpha}$ has absolute value r/R and angle $\theta-\alpha$.

Lengths and Transposes in the Complex Case

We return to linear algebra, and make the conversion from real to complex. By definition, *the complex vector space* $\mathbf{C}^n$ *contains all vectors x with n complex components*:

$$\textbf{Complex vector} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{with components} \quad x_j = a_j + ib_j.$$

Vectors x and y are still added component by component. Scalar multiplication cx is now done with complex numbers c. The vectors $v_1, \ldots, v_k$ are linearly *dependent* if some nontrivial combination gives $c_1v_1 + \ldots + c_kv_k = 0$; the c_j may now be complex. The unit coordinate vectors are still in $\mathbf{C}^n$; they are still independent; and they still form a basis. Therefore $\mathbf{C}^n$ is a complex vector space of dimension n.

In the new definition of length, each x_j^2 is replaced by its modulus $|x_j|^2$:

$$\textbf{Length squared} \qquad \|x\|^2 = |x_1|^2 + \cdots + |x_n|^2. \tag{4}$$

Example 2. $x = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $\|x\|^2 = 2$; $\quad y = \begin{bmatrix} 2+i \\ 2-4i \end{bmatrix}$ and $\|y\|^2 = 25$.

For real vectors there was a close connection between the length and the inner product: $\|x\|^2 = x^{\mathrm{T}}x$. This connection we want to preserve. The inner product must be modified to match the new definition of length, and we *conjugate the first vector in the inner product*. Replacing x by $\bar{x}$, ***the inner product becomes***

$$\textbf{Inner product} \qquad \bar{x}^{\mathrm{T}}y = \bar{x}_1y_1 + \cdots + \bar{x}_ny_n. \tag{5}$$

If we take the inner product of $x = (1+3i, 3i)$ with itself, we are back to $\|x\|^2$:

$$\textbf{Length squared} \qquad \bar{x}^{\mathrm{T}}x = \overline{(1+i)}(1+i) + \overline{(3i)}(3i) = 2+9 \quad \text{and} \quad \|x\|^2 = 11.$$

Note that $\bar{y}^{\mathrm{T}}x$ is different from $\bar{x}^{\mathrm{T}}y$; we have to watch the order of the vectors.

This leaves only one more change in notation, condensing two symbols into one. Instead of a bar for the conjugate and a T for the transpose, those are combined into the ***conjugate transpose***. For vectors and matrices, a superscript H (or a star) combines both operations. This matrix $\overline{A}^{\mathrm{T}} = A^{\mathrm{H}} = A^*$ is called "A Hermitian":

$$\textbf{“}A\textbf{ Hermitian”} \qquad A^{\mathrm{H}} = \overline{A}^{\mathrm{T}} \quad \text{has entries} \quad (A^{\mathrm{H}})_{ij} = \overline{A_{ji}}. \tag{6}$$

You have to listen closely to distinguish that name from the phrase "A is Hermitian," which means that A *equals* A^{H}. If A is an m by n matrix, then A^{H} is n by m:

$$\begin{array}{c}\textbf{Conjugate}\\ \textbf{transpose}\end{array} \qquad \begin{bmatrix} 2+i & 3i \\ 4-i & 5 \\ 0 & 0 \end{bmatrix}^{\mathrm{H}} = \begin{bmatrix} 2-i & 4+i & 0 \\ -3i & 5 & 0 \end{bmatrix}.$$

This symbol A^{H} gives official recognition to the fact that, with complex entries, it is very seldom that we want only the transpose of A. It is the *conjugate* transpose A^{H} that becomes appropriate, and x^{H} is the row vector $[\bar{x}_1 \ \cdots \ \bar{x}_n]$.

5N

1. The inner product of x and y is $x^{\mathrm{H}}y$. Orthogonal vectors have $x^{\mathrm{H}}y = 0$.
2. The squared length of x is $\|x\|^2 = x^{\mathrm{H}}x = |x_1|^2 + \cdots + |x_n|^2$.
3. Conjugating $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$ produces $(AB)^{\mathrm{H}} = B^{\mathrm{H}}A^{\mathrm{H}}$.

Hermitian Matrices

We spoke in earlier chapters about symmetric matrices: $A = A^{\mathrm{T}}$. With complex entries, this idea of symmetry has to be extended. The right generalization is not to matrices that equal their transpose, but to ***matrices that equal their conjugate transpose***. These are the Hermitian matrices, and a typical example is A:

$$\textbf{Hermitian matrix} \qquad A = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} = A^{\mathrm{H}}. \tag{7}$$

The diagonal entries must be real; they are unchanged by conjugation. Each off-diagonal entry is matched with its mirror image across the main diagonal, and $3-3i$ is the conjugate of $3+3i$. *In every case,* $a_{ij} = \overline{a_{ji}}$.

Our main goal is to establish three basic properties of Hermitian matrices. These properties apply equally well to symmetric matrices. *A real symmetric matrix is certainly Hermitian.* (For real matrices there is no difference between A^{T} and A^{H}.) **The eigenvalues of** A **are real**—as we now prove.

Property 1 If $A = A^{\mathrm{H}}$, then for all complex vectors x, the number $x^{\mathrm{H}}Ax$ is real.

Every entry of A contributes to $x^{\mathrm{H}}Ax$. Try the 2 by 2 case with $x = (u, v)$:

$$\begin{aligned} x^{\mathrm{H}}Ax &= \begin{bmatrix} \overline{u} & \overline{v} \end{bmatrix} \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ &= 2\overline{u}u + 5\overline{v}v + (3-3i)\overline{u}v + (3+3i)u\overline{v} \\ &= \textbf{real} + \textbf{real} + (\textbf{sum of complex conjugates}). \end{aligned}$$

For a proof in general. $(x^{\mathrm{H}}Ax)^{\mathrm{H}}$ is the conjugate of the 1 by 1 matrix $x^{\mathrm{H}}Ax$, but we actually get the same number back again: $(x^{\mathrm{H}}Ax)^{\mathrm{H}} = x^{\mathrm{H}}A^{\mathrm{H}}x^{\mathrm{HH}} = x^{\mathrm{H}}Ax$. So that number must be real.

Property 2 If $A = A^{\mathrm{H}}$, every eigenvalue is real.

Proof. Suppose $Ax = \lambda x$. ***The trick is to multiply by*** x^{H}: $x^{\mathrm{H}}Ax = \lambda x^{\mathrm{H}}x$. The left-hand side is real by Property 1, and the right-hand side $x^{\mathrm{H}}x = \|x\|^2$ is real and positive, because $x \neq 0$. Therefore $\lambda = x^{\mathrm{H}}Ax/x^{\mathrm{H}}x$ must be real. Our example has $\lambda = 8$ and $\lambda = -1$:

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 2-\lambda & 3-3i \\ 3+3i & 5-\lambda \end{vmatrix} = \lambda^2 - 7\lambda + 10 - |3-3i|^2 \\ &= \lambda^2 - 7\lambda - 8 = (\lambda - 8)(\lambda + 1). \end{aligned} \tag{8}$$

□

Note. This proof of real eigenvalues looks correct for any real matrix:

$$\textbf{False proof} \qquad Ax = \lambda x \quad \text{gives} \quad x^{\mathrm{T}}Ax = \lambda x^{\mathrm{T}}x, \quad \text{so} \quad \lambda = \frac{x^{\mathrm{T}}Ax}{x^{\mathrm{T}}x} \quad \text{is real.}$$

There must be a catch: *The eigenvector x might be complex.* It is when $A = A^{\mathrm{T}}$ that we can be sure λ and x stay real. More than that, *the eigenvectors are perpendicular*: $x^{\mathrm{T}}y = 0$ in the real symmetric case and $x^{\mathrm{H}}y = 0$ in the complex Hermitian case.

Property 3 Two eigenvectors of a real symmetric matrix or a Hermitian matrix, if they come from different eigenvalues, are orthogonal to one another.

The proof starts with $Ax = \lambda_1 x$, $Ay = \lambda_1 y$, and $A = A^{\mathrm{H}}$:

$$(\lambda_1 x)^{\mathrm{H}}y = (Ax)^{\mathrm{H}}y = x^{\mathrm{H}}Ay = x^{\mathrm{H}}(\lambda_2 y). \tag{9}$$

The outside numbers are $\lambda_1 x^{\mathrm{H}}y = \lambda_2 x^{\mathrm{H}}y$, since the λ's are real. Now wc use the assumption $\lambda_1 \neq \lambda_2$, *which forces the conclusion that* $x^{\mathrm{H}}y = 0$. In our example,

$$\begin{aligned} (A - 8I)x &= \begin{bmatrix} -6 & 3-i \\ 3+3i & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad x = \begin{bmatrix} 1 \\ 1+i \end{bmatrix} \\ (A + I)y &= \begin{bmatrix} 3 & 3-3i \\ 3+3i & 6 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad y = \begin{bmatrix} 1-i \\ -1 \end{bmatrix}. \end{aligned}$$

These two eigenvectors are orthogonal:

$$x^{\mathrm{H}}y = \begin{bmatrix} 1 & 1-i \end{bmatrix} \begin{bmatrix} 1-i \\ -1 \end{bmatrix} = 0.$$

Of course any multiples x/α and y/β are equally good as eigenvectors. MATLAB picks $\alpha = \|x\|$ and $\beta = \|y\|$, so that x/α and y/β are unit vectors; the eigenvectors are normalized to have length 1. They are now *orthonormal*. If these eigenvectors are chosen to be the columns of S, then we have $S^{-1}AS = \Lambda$ as always. ***The diagonalizing matrix can be chosen with orthonormal columns when*** $A = A^{\mathrm{H}}$.

In case A is real and symmetric, its eigenvalues are real by Property 2. Its unit eigenvectors are orthogonal by Property 3. Those eigenvectors are also real; they solve $(A - \lambda I)x = 0$. *These orthonormal eigenvectors go into an orthogonal matrix* Q, with $Q^{\mathrm{T}}Q = I$ and $Q^{\mathrm{T}} = Q^{-1}$. Then $S^{-1}AS = \Lambda$ becomes special—it is $Q^{-1}AQ = \Lambda$ or $A = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}$. We can state one of the great theorems of linear algebra:

5O A real symmetric matrix can be factored into $A = Q\Lambda Q^{\mathrm{T}}$. Its orthonormal eigenvectors are in the orthogonal matrix Q and its eigenvalues are in Λ.

In geometry or mechanics, this is the *principal axis theorem*. It gives the right choice of axes for an ellipse. Those axes are perpendicular, and they point along the eigenvectors of the corresponding matrix. (Section 6.2 connects symmetric matrices to n-dimensional ellipses.) In mechanics the eigenvectors give the principal directions, along which there is pure compression or pure tension—with no shear.

In mathematics the formula $A = Q\Lambda Q^{\mathrm{T}}$ is known as the *spectral theorem*. If we multiply columns by rows, the matrix A becomes a combination of one-dimensional projections—which are the special matrices xx^{T} of rank 1, multiplied by λ:

$$\begin{aligned} A = Q\Lambda Q^{\mathrm{T}} &= \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} - & x_1^{\mathrm{T}} & - \\ & \vdots & \\ - & x_n^{\mathrm{T}} & - \end{bmatrix} \\ &= \lambda_1 x_1 x_1^{\mathrm{T}} + \lambda_2 x_2 x_2^{\mathrm{T}} + \cdots + \lambda_n x_n x_n^{\mathrm{T}}. \end{aligned} \tag{10}$$

Our 2 by 2 example has eigenvalues 3 and 1:

Example 3. $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 3 \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} =$ **combination of two projections.**

The eigenvectors, with length scaled to 1, are

$$x_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \text{and} \qquad x_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Then the matrices on the right-hand side are $x_1 x_1^{\mathrm{T}}$ and $x_2 x_2^{\mathrm{T}}$—columns times rows—and they are projections onto the line through x_1 and the line through x_2.

All symmetric matrices are combinations of one-dimensional projections—which are symmetric matrices of rank 1.

Remark. If A is real and its eigenvalues happen to be real, then its eigenvectors are also real. They solve $(A-\lambda I)x = 0$ and can be computed by elimination. But they will not be orthogonal unless A is symmetric: $A = Q\Lambda Q^{\mathrm{T}}$ leads to $A^{\mathrm{T}} = A$.

If A is real, all complex eigenvalues come in conjugate pairs: $Ax = \lambda x$ and $A\bar{x} = \overline{\lambda}\bar{x}$. *If $a+ib$ is an eigenvalue of a real matrix, so is $a-ib$.* (If $A = A^{\mathrm{T}}$ then $b = 0$.)

Strictly speaking, the spectral theorem $A = Q\Lambda Q^{\mathrm{T}}$ has been proved only when the eigenvalues of A are distinct. Then there are certainly n independent eigenvectors, and A can be safely diagonalized. Nevertheless it is true (see Section 5.6) that *even with repeated eigenvalues, a symmetric matrix still has a complete set of orthonormal eigenvectors*. The extreme case is the identity matrix, which has $\lambda = 1$ repeated n times—and no shortage of eigenvectors.

To finish the complex case we need the analogue of a real orthogonal matrix—and you can guess what happens to the requirement $Q^{\mathrm{T}}Q = I$. The transpose will be replaced by the conjugate transpose. The condition will become $U^{\mathrm{H}}U = I$. The new letter U reflects the new name: ***A complex matrix with orthonormal columns is called a unitary matrix***.

Unitary Matrices

May we propose two analogies? ***A Hermitian (or symmetric) matrix can be compared to a real number. A unitary (or orthogonal) matrix can be compared to a number on the unit circle***—a complex number of absolute value 1. The λ's are real if $A^{\mathrm{H}} = A$, and they are on the unit circle if $U^{\mathrm{H}}U = I$. The eigenvectors can be scaled to unit length and made orthonormal.[6]

Those statements are not yet proved for unitary (including orthogonal) matrices. Therefore we go directly to the three properties of U that correspond to the earlier Properties 1–3 of A. Remember that U has orthonormal columns:

$$\textbf{Unitary matrix} \qquad U^{\mathrm{H}}U = I, \qquad UU^{\mathrm{H}} = I, \quad \text{and} \quad U^{\mathrm{H}} = U^{-1}.$$

This leads directly to Property $1'$, that multiplication by U has no effect on inner products, angles, or lengths. The proof is on one line, just as it was for Q:

Property $1'$ $(Ux)^{\mathrm{H}}(Uy) = x^{\mathrm{H}}U^{\mathrm{H}}Uy = x^{\mathrm{H}}y$ and lengths are preserved by U:

$$\textbf{Length unchanged} \qquad \|Ux\|^2 = x^{\mathrm{H}}U^{\mathrm{H}}Ux = \|x\|^2. \tag{11}$$

Property $2'$ Every eigenvalue of U has absolute value $|\lambda| = 1$.

This follows directly from $Ux = \lambda x$, by comparing the lengths of the two sides: $\|Ux\| = \|x\|$ by Property $1'$, and always $\|\lambda x\| = |\lambda|\|x\|$. Therefore $|\lambda| = 1$.

[6]Later we compare "skew-Hermitian" matrices with pure imaginary numbers, and "normal" matrices with all complex numbers $a+ib$. A nonnormal matrix without orthogonal eigenvectors belongs to none of these classes, and is outside the whole analogy.

Property 3′ Eigenvectors corresponding to different eigenvalues are orthonormal.

Start with $Ux = \lambda_1 x$ and $Uy = \lambda_2 y$, and take inner products by Property 1′:

$$x^{\mathrm{H}}y = (Ux)^{\mathrm{H}}(Uy) = (\lambda_1 x)^{\mathrm{H}}(\lambda_2 y) = \overline{\lambda}_1\lambda_2 x^{\mathrm{H}}y.$$

Comparing the left to the right, $\overline{\lambda}_1\lambda_2 = 1$ or $x^{\mathrm{H}}y = 0$. But Property 2′ is $\overline{\lambda}_1\lambda_1 = 1$, so we cannot also have $\overline{\lambda}_1\lambda_2 = 1$. Thus $x^{\mathrm{H}}y = 0$ and the eigenvectors are orthogonal.

Example 4. $U = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}$ has eigenvalues e^{it} and e^{-it}.

The orthogonal eigenvectors are $x = (1, -i)$ and $y = (1, i)$. (Remember to take conjugates in $x^{\mathrm{H}}y = 1 + i^2 = 0$.) After division by $\sqrt{2}$ they are orthonormal.

Here is the most important *unitary matrix* by far.

Example 5. $U = \dfrac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdot & 1 \\ 1 & w & \cdot & w^{n-1} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & w^{n-1} & \cdot & w^{(n-1)^2} \end{bmatrix} = \dfrac{\textbf{Fourier matrix}}{\sqrt{n}}.$

The complex number w is on the unit circle at the angle $\theta = 2\pi/n$. It equals $e^{2\pi i/n}$. Its powers are spaced evenly around the circle. That spacing assures that the sum of all n powers of w—all the nth roots of 1—is zero. Algebraically, the sum $1 + w + \cdots + w^{n-1}$ is $(w^n - 1)/(w - 1)$. And $w^n - 1$ is zero!

$$\text{row 1 of } U^{\mathrm{H}} \text{ times column 2 of } U \text{ is } \frac{1}{n}(1 + w + w^2 + \cdots + w^{n-1}) = \frac{w^n - 1}{w - 1} = 0.$$

$$\text{row } i \text{ of } U^{\mathrm{H}} \text{ times column } j \text{ of } U \text{ is } \frac{1}{n}(1 + W + W^2 + \cdots + W^{n-1}) = \frac{W^n - 1}{W - 1} = 0.$$

In the second case, $W = w^{j-i}$. Every entry of the original F has absolute value 1. The factor $\sqrt{n}$ shrinks the columns of U into unit vectors. ***The fundamental identity of the finite Fourier transform is*** $U^{\mathrm{H}}U = I$.

Thus U is a unitary matrix. Its inverse looks the same except that w is replaced by $w^{-1} = e^{-i\theta} = \overline{w}$. Since U is unitary, its inverse is found by transposing (which changes nothing) and conjugating (which changes w to $\overline{w}$). The inverse of this U is $\overline{U}$. Ux can be computed quickly by the ***Fast Fourier Transform*** as found in Section 3.5.

By Property 1′ of unitary matrices, the length of a vector x is the same as the length of Ux. The energy in state space equals the energy in transform space. The energy is the sum of $|x_j|^2$, and it is also the sum of the energies in the separate frequencies. The vector $x = (1, 0, \ldots, 0)$ contains equal amounts of every frequency component, and its Discrete Fourier Transform $Ux = (1, 1, \ldots, 1)/\sqrt{n}$ also has length 1.

Example 6.

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

This is an orthogonal matrix, so by Property 3′ it must have orthogonal eigenvectors. They are the columns of the Fourier matrix! Its eigenvalues must have absolute value 1. They are the numbers $1, w, \dots, w^{n-1}$ (or $1, i, i^2, i^3$ in this 4 by 4 ease). It is a real matrix, but its eigenvalues and eigenvectors are complex.

One final note, Skew-Hermitian matrices satisfy $K^{\mathrm{H}} = -K$, just as skew-symmetric matrices satisfy $K^{\mathrm{T}} = -K$. Their properties follow immediately from their close link to Hermitian matrices:

If** A **is Hermitian then** $K = iA$ **is skew-Hermitian.

The eigenvalues of K are purely imaginary instead of purely real; we multiply i. The eigenvectors are not changed. The Hermitian example on the previous pages would lead to

$$K = iA = \begin{bmatrix} 2i & 3+3i \\ -3+3i & 5i \end{bmatrix} = -K^{\mathrm{H}}.$$

The diagonal entries are multiples of i (allowing zero). The eigenvalues are $8i$ and $-i$. The eigenvectors are still orthogonal, and we still have $K = U\Lambda U^{\mathrm{H}}$—with a unitary U instead of a real orthogonal Q, and with $8i$ and $-i$ on the diagonal of Λ.

This section is summarized by a table of parallels between real and complex.

Real versus Complex

$\mathbf{R}^n$ (n real components)	$\leftrightarrow$	$\mathbf{C}^n$ (n complex components)
length: $\|x\|^2 = x_1^2 + \cdots + x_n^2$	$\leftrightarrow$	length: $\|x\|^2 = \lvert x_1\rvert^2 + \cdots + \lvert x_n\rvert^2$
transpose: $A_{ij}^{\mathrm{T}} = A_{ji}$	$\leftrightarrow$	Hermitian transpose: $A_{ij}^{\mathrm{H}} = \overline{A_{ji}}$
$(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$	$\leftrightarrow$	$(AB)^{\mathrm{H}} = B^{\mathrm{H}}A^{\mathrm{H}}$
inner product: $x^{\mathrm{T}}y = x_1y_1 + \cdots + x_ny_n$	$\leftrightarrow$	inner product: $x^{\mathrm{H}}y = \bar{x}_1y_1 + \cdots + \bar{x}_ny_n$
$(Ax)^{\mathrm{T}}y = x^{\mathrm{T}}(A^{\mathrm{T}}y)$	$\leftrightarrow$	$(Ax)^{\mathrm{H}}y = x^{\mathrm{H}}(A^{\mathrm{H}}y)$
orthogonality: $x^{\mathrm{T}}y = 0$	$\leftrightarrow$	orthogonality: $x^{\mathrm{H}}y = 0$
symmetric matrices: $A^{\mathrm{T}} = A$	$\leftrightarrow$	Hermitian matrices: $A^{\mathrm{H}} = A$
$A = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}$ (real Λ)	$\leftrightarrow$	$A = U\Lambda U^{-1} = U\Lambda U^{\mathrm{H}}$ (real Λ)
skew-symmetric $K^{\mathrm{T}} = -K$	$\leftrightarrow$	skew-Hermitian $K^{\mathrm{H}} = -K$
orthogonal $Q^{\mathrm{T}}Q = I$ or $Q^{\mathrm{T}} = Q^{-1}$	$\leftrightarrow$	unitary $U^{\mathrm{H}}U = I$ or $U^{\mathrm{H}} = U^{-1}$
$(Qx)^{\mathrm{T}}(Qy) = x^{\mathrm{T}}y$ and $\|Qx\| = \|x\|$	$\leftrightarrow$	$(Ux)^{\mathrm{H}}(Uy) = x^{\mathrm{H}}y$ and $\|Ux\| = \|x\|$

The columns, rows, and eigenvectors of Q and U are orthonormal, and every $|\lambda| = 1$

Problem Set 5.5

1. For the complex numbers $3+4i$ and $1-i$,

(a) find their positions in the complex plane.

(b) find their sum and product.

(c) find their conjugates and their absolute values.

Do the original numbers lie inside or outside the unit circle?

2. What can you say about

(a) the sum of a complex number and its conjugate?

(b) the conjugate of a number on the unit circle?

(c) the product of two numbers on the unit circle?

(d) the sum of two numbers on the unit circle?

3. If $x=2+i$ and $y=1+3i$, find $\bar{x}$, $x\bar{x}$, $1/x$, and x/y. Check that the absolute value $|xy|$ equals $|x|$ times $|y|$, and the absolute value $|1/x|$ equals 1 divided by $|x|$.

4. Find a and b for the complex numbers $a+ib$ at the angles $\theta = 30^\circ, 60^\circ, 90^\circ$ on the unit circle. Verify by direct multiplication that the square of the first is the second, and the cube of the first is the third.

5. (a) If $x = re^{i\theta}$ what are x^2, x^{-1}, and $\bar{x}$ in polar coordinates? Where are the complex numbers that have $x^{-1} = \bar{x}$?

(b) At $t=0$, the complex number $e^{(-1+i)t}$ equals one. Sketch its path in the complex plane as t increases from 0 to 2π.

6. Find the lengths and the inner product of

$$x = \begin{bmatrix} 2-4i \\ 4i \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} 2+4i \\ 4i \end{bmatrix}.$$

7. Write out the matrix A^{H} and compute $C = A^{\mathrm{H}}A$ if

$$A = \begin{bmatrix} 1 & i & 0 \\ i & 0 & 1 \end{bmatrix}.$$

What is the relation between C and C^{H}? Does it hold whenever C is constructed from some $A^{\mathrm{H}}A$?

8. (a) With the preceding A, use elimination to solve $Ax=0$.

(b) Show that the nullspace you just computed is orthogonal to $\boldsymbol{C}(A^{\mathrm{H}})$ *and not to the usual row space* $\boldsymbol{C}(A^{\mathrm{T}})$. The four fundamental spaces in the complex case are $\boldsymbol{N}(A)$ and $\boldsymbol{C}(A)$ as before, and then $\boldsymbol{N}(A^{\mathrm{H}})$ and $\boldsymbol{C}(A^{\mathrm{H}})$.

9. (a) How is the determinant of A^{H} related to the determinant of A?

(b) Prove that the determinant of any Hermitian matrix is real.

10. (a) How many degrees of freedom are there in a real symmetric matrix, a real diagonal matrix, and a real orthogonal matrix? (The first answer is the sum of the other two, because $A = Q\Lambda Q^{\mathrm{T}}$.)

(b) Show that 3 by 3 Hermitian matrices A and also unitary U have 9 real degrees of freedom (columns of U can be multiplied by any $e^{i\theta}$).

11. Write P, Q and R in the form $\lambda_1 x_1 x_1^{\mathrm{H}} + \lambda_2 x_2 x_2^{\mathrm{H}}$ of the spectral theorem:

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \qquad Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad R = \begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}.$$

12. Give a reason if true or a counterexample if false:

(a) If A is Hermitian, then $A + iI$ is invertible.

(b) If Q is orthogonal. then $Q + \frac{1}{2}I$ is invertible.

(c) If A is real, then $A + iI$ is invertible.

13. Suppose A is a symmetric 3 by 3 matrix with eigenvalues 0, 1, 2.

(a) What properties can be guaranteed for the corresponding unit eigenvectors u, v, w?

(b) In terms of u, v, w, describe the nullspace, left nullspace, row space and column space of A.

(c) Find a vector x that satisfies $Ax = v + w$. Is x unique?

(d) Under what conditions on b does $Ax = b$ have a solution?

(e) If u, v, w are the columns of S, what are S^{-1} and $S^{-1}AS$?

14. In the list below, which classes of matrices contain A and which contain B?

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad B = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

*Orthogonal, invertible, projection, permutation, Hermitian, rank-*1*, diagonalizable, Markov*. Find the eigenvalues of A and B.

15. What is the dimension of the space S of all n by n real symmetric matrices? The spectral theorem says that every symmetric matrix is a combination of n projection matrices. Since the dimension exceeds n, how is this difference explained?

16. Write one significant fact about the eigenvalues of each of the following.

(a) A real symmetric matrix.

(b) A stable matrix: all solutions to $du/dt = Au$ approach zero.

(c) An orthogonal matrix.

(d) A Markov matrix.

(e) A defective matrix (nondiagonalizable).

(f) A singular matrix.

17. Show that if U and V are unitary, so is UV. Use the criterion $U^{\mathrm{H}}U = I$.

18. Show that a unitary matrix has $|\det U| = 1$, but possibly $\det U$ is different from $\det U^{\mathrm{H}}$. Describe all 2 by 2 matrices that are unitary.

19. Find a third column so that U is unitary. How much freedom in column 3?

$$U = \begin{bmatrix} 1/\sqrt{3} & i/\sqrt{2} & \\ 1/\sqrt{3} & 0 & \\ i/\sqrt{3} & 1/\sqrt{2} & \end{bmatrix}.$$

20. Diagonalize the 2 by 2 skew-Hermitian matrix $K = \left[\begin{smallmatrix} i & i \\ i & i \end{smallmatrix}\right]$, whose entries are all $\sqrt{-1}$. Compute $e^{Kt} = Se^{\Lambda t}S^{-1}$, and verify that e^{Kt} is unitary. What is the derivative of e^{Kt} at $t = 0$?

21. Describe all 3 by 3 matrices that are simultaneously Hermitian, unitary, and diagonal. How many are there?

22. Every matrix Z can be split into a Hermitian and a skew-Hermitian part, $Z = A + K$, just as a complex number z is split into $a + ib$, The real part of z is half of $z + \bar{z}$, and the "real part" of Z is half of $Z + Z^{\mathrm{H}}$. Find a similar formula for the "imaginary part" K, and split these matrices into $A + K$:

$$Z = \begin{bmatrix} 3+i & 4+2i \\ 0 & 5 \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} i & i \\ -i & i \end{bmatrix}.$$

23. Show that the columns of the 4 by 4 Fourier matrix F in Example 5 are eigenvectors of the permutation matrix P in Example 6.

24. For the permutation of Example 6, write out the *circulant matrix* $C = c_0I + c_1P + c_2P^2 + c_3P^3$. (Its eigenvector matrix is again the Fourier matrix.) Write out also the four components of the matrix-vector product Cx, which is the *convolution* of $c = (c_0, c_1, c_2, c_3)$ and $x = (x_0, x_1, x_2, x_3)$.

25. For a circulant $C = F\Lambda F^{-1}$, why is it faster to multiply by F^{-1}, then Λ, then F (the convolution rule), than to multiply directly by C?

26. Find the lengths of $u = (1+i, 1-i, 1+2i)$ and $v = (i, i, i)$. Also find $u^{\mathrm{H}}v$ and $v^{\mathrm{H}}u$.

27. Prove that $A^{\mathrm{H}}A$ is always a Hermitian matrix, Compute $A^{\mathrm{H}}A$ and AA^{H}:

$$A = \begin{bmatrix} i & 1 & i \\ 1 & i & i \end{bmatrix}.$$

28. If $Az = 0$, then $A^{\mathrm{H}}Az = 0$. If $A^{\mathrm{H}}Az = 0$, multiply by z^{H} to prove that $Az = 0$. The nullspaces of A and $A^{\mathrm{H}}A$ are ____. $A^{\mathrm{H}}A$ is an invertible Hermitian matrix when the nullspace of A contains only $z =$ ____.

29. When you multiply a Hermitian matrix by a real number c, is cA still Hermitian? If $c = i$, show that iA is skew-Hermitian. The 3 by 3 Hermitian matrices are a subspace, provided that the "scalars" are real numbers.

30. Which classes of matrices does P belong to: orthogonal, invertible, Hermitian, unitary, factorizable into LU, factorizable into QR?

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

31. Compute P^2, P^3, and P^{100} in Problem 30. What are the eigenvalues of P?

32. Find the unit eigenvectors of P in Problem 30, and put them into the columns of a unitary matrix U. What property of P makes these eigenvectors orthogonal?

33. Write down the 3 by 3 *circulant matrix* $C = 2I + 5P + 4P^2$. It has the same eigenvectors as P in Problem 30. Find its eigenvalues.

34. If U is unitary and Q is a real orthogonal matrix, show that U^{-1} is unitary and also UQ is unitary. Start from $U^{\mathrm{H}}U = I$ and $Q^{\mathrm{T}}Q = I$.

35. Diagonalize A (real λ's) and K (imaginary λ's) to reach $U\Lambda U^{\mathrm{H}}$:

$$A = \begin{bmatrix} 0 & 1-i \\ i+1 & 1 \end{bmatrix} \qquad K = \begin{bmatrix} 0 & -1+i \\ 1+i & i \end{bmatrix}$$

36. Diagonalize this orthogonal matrix to reach $Q = U\Lambda U^{\mathrm{H}}$. Now all λ's are ____:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

37. Diagonalize this unitary matrix V to reach $V = U\Lambda U^{\mathrm{H}}$. Again all $|\lambda| = 1$:

$$V = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}.$$

38. If $v_1, \dots, v_n$ is an orthonormal basis for $\mathbf{C}^n$, the matrix with those columns is a ____ matrix. Show that any vector z equals $(v_1^{\mathrm{H}} z)v_1 + \cdots + (v_n^{\mathrm{H}} z)v_n$.

39. The functions e^{-ix} and e^{-ix} are orthogonal on the interval $0 \le x \le 2\pi$ because their *complex* inner product is $\int_0^{2\pi}$ ____ $= 0$.

40. The vectors $v = (1, i, 1)$, $w = (i, 1, 0)$ and $z =$ ____ are an orthogonal basis for ____.

41. If $A = R + iS$ is a Hermitian matrix, are the real matrices R and S symmetric?

42. The (complex) dimension of $\mathbf{C}^n$ is ____. Find a nonreal basis for $\mathbf{C}^n$.

43. Describe all 1 by 1 matrices that are Hermitian and also unitary. Do the same for 2 by 2 matrices.

44. How are the eigenvalues of A^{H} (square matrix) related to the eigenvalues of A?

45. If $u^{\mathrm{H}} u = 1$, show that $I - 2uu^{\mathrm{H}}$ is Hermitian and also unitary. The rank-1 matrix uu^{H} is the projection onto what line in $\mathbf{C}^n$?

46. If $A + iB$ is a unitary matrix (A and B are real), show that $Q = \left[\begin{smallmatrix} A & -B \\ B & A \end{smallmatrix}\right]$ is an orthogonal matrix.

47. If $A + iB$ is a Hermitian matrix (A and B are real), show that $\left[\begin{smallmatrix} A & -B \\ B & A \end{smallmatrix}\right]$ is symmetric.

48. Prove that the inverse of a Hermitian matrix is again a Hermitian matrix.

49. Diagonalize this matrix by constructing its eigenvalue matrix Λ and its eigenvector matrix S:

$$A = \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix} = A^{\mathrm{H}}.$$

50. A matrix with orthonormal eigenvectors has the form $A = U\Lambda U^{-1} = U\Lambda U^{\mathrm{H}}$. *Prove that* $AA^{\mathrm{H}} = A^{\mathrm{H}}A$. These are exactly the *normal matrices*.

5.6 Similarity Transformations

Virtually every step in this chapter has involved the combination $S^{-1}AS$. The eigenvectors of A went into the columns of S, and that made $S^{-1}AS$ a diagonal matrix (called Λ). When A was symmetric, we wrote Q instead of S, choosing the eigenvectors to be orthonormal. In the complex case, when A is Hermitian we write U—it is still the matrix of eigenvectors. Now we look at all combinations $M^{-1}AM$—*formed with any invertible* M *on the right and its inverse on the left*. The invertible eigenvector matrix S may fail to exist (the defective case), or we may not know it, or we may not want to use it.

First a new word: ***The matrices*** A ***and*** $M^{-1}AM$ ***are "similar".*** Going from one to the other is a ***similarity transformation***. It is the natural step for differential equations

Chapter 6

Positive Definite Matrices

6.1 Minima, Maxima, and Saddle Points

Up to now, we have hardly thought about **the signs of the eigenvalues**. We couldn't ask whether λ was positive before it was known to be real. Chapter 5 established that every symmetric matrix has real eigenvalues. Now we will find a test that can be applied directly to A, without computing its eigenvalues, which will ***guarantee that all those eigenvalues are positive***. The test brings together three of the most basic ideas in the book—*pivots*, *determinants*, and *eigenvalues*.

The signs of the eigenvalues are often crucial. For stability in differential equations, we needed negative eigenvalues so that $e^{\lambda t}$ would decay. The new and highly important problem is to recognize a ***minimum point***. This arises throughout science and engineering and every problem of optimization. The mathematical problem is to move the second derivative test $F'' > 0$ into n dimensions. Here are two examples:

$$F(x,y) = 7 + 2(x+y)^2 - y\sin y - x^3 \qquad f(x,y) = 2x^2 + 4xy + y^2.$$

Does either $F(x,y)$ or $f(x,y)$ have a minimum at the point $x = y = 0$?

***Remark* 3.** The *zero-order terms* $F(0,0) = 7$ and $f(0,0) = 0$ have no effect on the answer. They simply raise or lower the graphs of F and f.

***Remark* 4.** The *linear terms* give a necessary condition: To have any chance of a minimum, the first derivatives must vanish at $x = y = 0$:

$$\frac{\partial F}{\partial x} = 4(x+y) - 3x^2 = 0 \qquad \text{and} \qquad \frac{\partial F}{\partial y} = 4(x+y) - y\cos y - \sin y = 0$$

$$\frac{\partial f}{\partial x} = 4x + 4y = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 4x + 2y = 0. \quad \textit{All zero}.$$

Thus $(x,y) = (0,0)$ is a stationary point for both functions. The surface $z = F(x,y)$ is tangent to the horizontal plane $z = 7$, and the surface $z = f(x,y)$ is tangent to the plane $z = 0$. The question is whether the graphs go *above those planes or not*, as we move away from the tangency point $x = y = 0$.

Remark* 5. *The second derivatives at* (0,0) *are decisive:

$$\frac{\partial^2 F}{\partial x^2} = 4 - 6x = 4 \qquad \frac{\partial^2 f}{\partial x^2} = 4$$
$$\frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2 F}{\partial y \partial x} = 4 \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 4$$
$$\frac{\partial^2 F}{\partial y^2} = 4 + y \sin y - 2\cos y = 2 \qquad \frac{\partial^2 f}{\partial y^2} = 2.$$

These second derivatives 4, 4, 2 contain the answer. Since they are the same for F and f, they must contain the same answer. The two functions behave in exactly the same way near the origin. ***F has a minimum if and only if f has a minimum***. I am going to show that those functions don't!

***Remark* 6.** The *higher-degree terms* in F have no effect on the question of a *local* minimum, but they can prevent it from being a *global* minimum. In our example the term $-x^3$ must sooner or later pull F toward $-\infty$. For $f(x,y)$, with no higher terms, all the action is at $(0,0)$.

Every quadratic form $f = ax^2 + 2bxy + cy^2$ has a stationary point at the origin, where $\partial f/\partial x = \partial f/\partial y = 0$. A local minimum would also be a global minimum, The surface $z = f(x,y)$ will then be shaped like a bowl, resting on the origin (Figure 6.1). If the stationary point of F is at $x = \alpha$, $y = \beta$, the only change would be to use the second derivatives at α, β:

Quadratic part of F
$$f(x,y) = \frac{x^2}{2}\frac{\partial^2 F}{\partial x^2}(\alpha,\beta) + xy\frac{\partial^2 F}{\partial x \partial y}(\alpha,\beta) + \frac{y^2}{2}\frac{\partial^2 F}{\partial y^2}(\alpha,\beta). \tag{1}$$

This $f(x,y)$ behaves near $(0,0)$ in the same way that $F(x,y)$ behaves near (α,β).

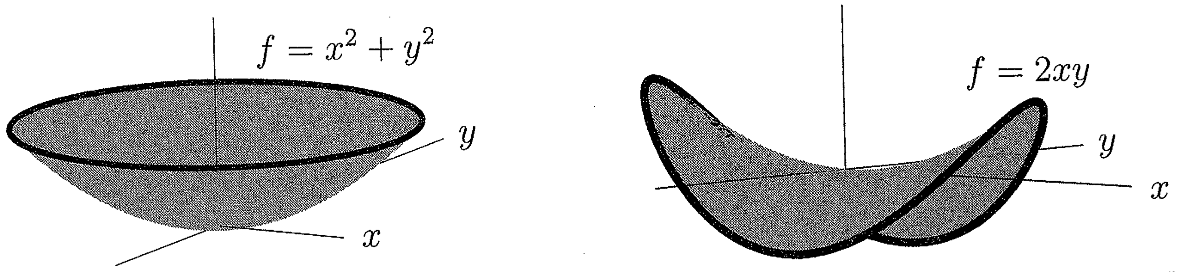

Figure 6.1: A bowl and a saddle: Definite $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and indefinite $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

The third derivatives are drawn into the problem when the second derivatives fail to give a definite decision. That happens when the quadratic part is singular. For a true minimum, f is allowed to vanish *only* at $x = y = 0$. When $f(x,y)$ is strictly positive at all other points (the bowl goes up), it is called ***positive definite***.

Definite versus Indefinite: Bowl versus Saddle

The problem comes down to this: For a function of two variables x and y, what is the correct replacement for the condition $\partial^2 F/\partial x^2 > 0$? With only one variable, the sign of the second derivative decides between a minimum or a maximum. Now we have three second derivatives: F_{xx}, $F_{xy} = F_{yx}$, and F_{yy}. These three numbers (like 4, 4, 2) must determine whether or not F (as well as f) has a minimum.

***What conditions on** a, b, **and** c **ensure that the quadratic** $f(x,y) = ax^2 + 2bxy + cy^2$ **is positive definite**?* One necessary condition is easy:

(i) *If $ax^2 + 2bxy + cy^2$ is positive definite, then necessarily $a > 0$.*

We look at $x = 1$, $y = 0$, where $ax^2 + 2bxy + cy^2$ is equal to a. This must be positive. Translating back to F, that means that $\partial^2 F/\partial x^2 > 0$. The graph must go up in the x direction. Similarly, fix $x = 0$ and look in the y direction where $f(0,y) = cy^2$:

(ii) *If $f(x,y)$ is positive definite, then necessarily $c > 0$.*

Do these conditions $a > 0$ and $c > 0$ guarantee that $f(x,y)$ is always positive? The answer is **no**. A large cross term $2bxy$ can pull the graph below zero.

Example 1. $f(x,y) = x^2 - 10xy + y^2$. Here $a = 1$ and $c = 1$ are both positive. But f is not positive definite, because $f(1,1) = -8$. The conditions $a > 0$ and $c > 0$ ensure that $f(x,y)$ is positive on the x and y axes. But this function is negative on the line $x = y$, because $b = -10$ overwhelms a and c.

Example 2. In our original f the coefficient $2b = 4$ was positive. Does this ensure a minimum? Again the answer is **no**; the sign of b is of no importance! *Even though its second derivatives are positive,* $2x^2 + 4xy + y^2$ *is not positive definite.* ***Neither F nor f has a minimum at*** $(0,0)$ ***because*** $f(1,-1) = 2 - 4 + 1 = -1$.

It is the size of** b, **compared to** a **and** c, **that must be controlled. We now want a necessary and sufficient condition for positive definiteness. The simplest technique is to complete the square:

Express $f(x,y)$ **using squares**

$$f = ax^2 + 2bxy + cy^2 = a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2. \quad (2)$$

The first term on the right is never negative, when the square is multiplied by $a > 0$. But this square can be zero, and the second term must then be positive. That term has coefficient $(ac - b^2)/a$. The last requirement for positive definiteness is that this coefficient must be positive:

(iii) If $ax^2 + 2bxy + cy^2$ stays positive, then necessarily $ac > b^2$.

Test for a minimum: The conditions $a > 0$ and $ac > b^2$ are just right. They guarantee $c > 0$. The right side of (2) is positive, and we have found a minimum:

6A $ax^2+2bxy+cy^2$ is positive definite if and only if $a>0$ and $ac>b^2$. Any $f(x,y)$ has a minimum at a point where $\partial F/\partial x=\partial F/\partial y=0$ with

$$\frac{\partial F^2}{\partial x^2}>0 \qquad \text{and} \qquad \left[\frac{\partial F^2}{\partial x^2}\right]\left[\frac{\partial F^2}{\partial y^2}\right]>\left[\frac{\partial F^2}{\partial x\partial y}\right]^2. \tag{3}$$

Test for a maximum: Since f has a maximum whenever $-f$ has a minimum, we just reverse the signs of a, b, and c. This actually leaves $ac>b^2$ unchanged: The quadratic form is ***negative definite*** if and only if $a<0$ and $ac>b^2$. The same change applies for a maximum of $F(x,y)$.

Singular case $ac=b^2$**:** The second term in equation (2) disappears to leave only the first square—which is either ***positive semidefinite***, when $a>0$, or ***negative semidefinite***, when $a<0$. The prefix *semi* allows the possibility that f can equal zero, as it will at the point $x=b$, $y=-a$. The surface $z=f(x,y)$ degenerates from a bowl into a valley. For $f=(x+y)^2$, the valley runs along the line $x+y=0$.

Saddle Point $ac<b^2$**:** In one dimension, $F(x)$ has a minimum or a maximum, or $F''=0$. In two dimensions, a very important possibility still remains: *The combination* $ac-b^2$ *may be negative*. This occurred in both examples, when b dominated a and c. It also occurs if a and c have opposite signs. Then two directions give opposite results—in one direction f increases, in the other it decreases. It is useful to consider two special cases:

$$\textbf{Saddle points at } (0,0) \qquad f_1=2xy \quad \text{and} \quad f_2=x^2-y^2 \quad \text{and} \quad ac-b^2=-1.$$

In the first, $b=1$ dominates $a=c=0$. In the second, $a=1$ and $c=-1$ have opposite sign. The saddles $2xy$ and x^2-y^2 are practically the same; if we turn one through 45° we get the other. They are also hard to draw.

These quadratic forms are ***indefinite***, because they can take either sign. So we have a stationary point that is neither a maximum or a minimum. It is called a ***saddle point***. The surface $z=x^2-y^2$ goes down in the direction of the y axis, where the legs fit (if you still ride a horse). In case you switched to a car, think of a road going over a mountain pass. The top of the pass is a minimum as you look along the range of mountains, but it is a maximum as you go along the road.

Higher Dimensions: Linear Algebra

Calculus would be enough to find our conditions $F_{xx}>0$ and $F_{xx}F_{yy}>F_{xy}^2$ for a minimum. But linear algebra is ready to do more, because the second derivatives fit into a symmetric matrix A. The terms ax^2 and cy^2 appear *on the diagonal*. The cross derivative $2bxy$ is split between the same entry b above and below. A quadratic $f(x,y)$ comes directly from a symmetric 2 by 2 matrix!

$$x^{\mathrm{T}}Ax \textbf{ in } \mathbf{R}^2 \qquad ax^2+2bxy+cy^2=\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}. \tag{4}$$

This identity (please multiply it out) is the key to the whole chapter. It generalizes immediately to n dimensions, and it is a perfect shorthand for studying maxima and minima. When the variables are $x_1,\dots,x_n$, they go into a column vector x. ***For any symmetric matrix** A**, the product** $x^{\mathrm{T}}Ax$ **is a pure quadratic form*** $f(x_1,\dots,x_n)$:

$$x^{\mathrm{T}}Ax \textbf{ in } \mathbf{R}^n \qquad \begin{bmatrix} x_1 & x_2 & \cdot & x_n \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix}=\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j. \tag{5}$$

The diagonal entries a_{11} to a_{nn} multiply x_1^2 to x_n^2. The pair $a_{ij}=a_{ji}$ combines into $2a_{ij}x_ix_j$. Then $f=a_{11}x_1^2+2a_{12}x_1x_2+\cdots+a_{nn}x_n^2$.

There are no higher-order terms or lower-order terms—only second-order. The function is zero at $x=(0,\dots,0)$, and its first derivatives are zero. The tangent is flat; this is a stationary point. We have to decide if $x=0$ is a minimum or a maximum or a saddle point of the function $f=x^{\mathrm{T}}Ax$.

Example 3. $f=2x^2+4xy+y^2$ and $A=\begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix} \rightarrow$ *saddle point*.

Example 4. $f=2xy$ and $A=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \rightarrow$ *saddle point*.

Example 5. A is 3 by 3 for $2x_1^2-2x_1x_2+2x_2^2-2x_2x_3+2x_3^2$:

$$f=\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \textit{minimum at } (0,0,0).$$

Any function $F(x_1,\dots,x_n)$ is approached in the same way. At a stationary point all first derivatives are zero. A is the **"second derivative matrix"** with entries $a_{ij}=\partial^2F/\partial x_i\partial x_j$. This automatically equals $a_{ji}=\partial^2F/\partial x_j\partial x_i$, so A is symmetric. **Then** F ***has a minimum when the pure quadratic*** $x^{\mathrm{T}}Ax$ ***is positive definite***. These second-order terms control F near the stationary point:

$$\textbf{Taylor series} \qquad F(x)=F(0)+x^{\mathrm{T}}(\text{grad } F)+\frac{1}{2}x^{\mathrm{T}}Ax+\text{higher order terms}. \tag{6}$$

At a stationary point, grad $F=(\partial F/\partial x_1,\dots,\partial F/\partial x_n)$ is a vector of zeros. The second derivatives in $x^{\mathrm{T}}Ax$ take the graph up or down (or saddle). If the stationary point is at x_0

instead of 0, $F(x)$ and all derivatives are computed at x_0. Then x changes to $x - x_0$ on the right-hand side.

The next section contains the tests to decide whether $x^{\mathrm{T}}Ax$ is positive (the bowl goes up from $x = 0$). Equivalently, **the tests decide whether the matrix A is positive definite**—which is the main goal of the chapter.

Problem Set 6.1

1. The quadratic $f = x^2 + 4xy + 2y^2$ has a saddle point at the origin, despite the fact that its coefficients are positive. Write f as a *difference of two squares*.

2. Decide for or against the positive definiteness of these matrices, and write out the corresponding $f = x^{\mathrm{T}}Ax$:

$$\text{(a)}\ \begin{bmatrix} 1 & 3 \\ 3 & 5 \end{bmatrix}. \qquad \text{(b)}\ \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \qquad \text{(c)}\ \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}. \qquad \text{(d)}\ \begin{bmatrix} -1 & 2 \\ 2 & -8 \end{bmatrix}.$$

The determinant in (b) is zero; along what line is $f(x,y) = 0$?

3. If a 2 by 2 symmetric matrix passes the tests $a > 0$, $ac > b^2$, solve the quadratic equation $\det(A - \lambda I) = 0$ and show that both eigenvalues are positive.

4. Decide between a minimum, maximum, or saddle point for the following functions.

(a) $F = -1 + 4(e^x - x) - 5x\sin y + 6y^2$ at the point $x = y = 0$.

(b) $F = (x^2 - 2x)\cos y$, with stationary point at $x = 1$, $y = \pi$.

5. (a) For which numbers b is the matrix $A = \left[\begin{smallmatrix} 1 & b \\ b & 9 \end{smallmatrix}\right]$ positive definite?

(b) Factor $A = LDL^{\mathrm{T}}$ when b is in the range for positive definiteness.

(c) Find the minimum value of $\frac{1}{2}(x^2 + 2bxy + 9y^2) - y$ for b in this range.

(d) What is the minimum if $b = 3$?

6. Suppose the positive coefficients a and c dominate b in the sense that $a + c > 2b$. Find an example that has $ac < b^2$, so the matrix is not positive definite.

7. (a) What 3 by 3 symmetric matrices A_1 and A_2 correspond to f_1 and f_2?

$$\begin{aligned} f_1 &= x_1^2 + x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 + 2x_2x_3 \\ f_2 &= x_1^2 + 2x_2^2 + 11x_3^2 - 2x_1x_2 - 2x_1x_3 - 4x_2x_3. \end{aligned}$$

(b) Show that f_1 is a *single* perfect square and not positive definite. Where is f_1 equal to 0?

(c) Factor A_2 into LL^{T}, Write $f_2 = x^{\mathrm{T}}A_2x$ as a sum of three squares.

8. If $A = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$ is positive definite, test $A^{-1} = \left[\begin{smallmatrix} p & q \\ q & r \end{smallmatrix}\right]$ for positive definiteness.

9. The quadratic $f(x_1,x_2) = 3(x_1+2x_2)^2 + 4x_2^2$ is positive. Find its matrix A, factor it into LDL^{T}, and connect the entries in D and L to 3, 2, 4 in f.

10. If $R = \left[\begin{smallmatrix} p & q \\ q & r \end{smallmatrix}\right]$, write out R^2 and check that it is positive definite unless R is singular.

11. (a) If $A = \left[\begin{smallmatrix} a & b \\ \bar{b} & c \end{smallmatrix}\right]$ is Hermitian (*complex* b), find its pivots and determinant.

(b) Complete the square for $x^{\mathrm{H}}Ax$. Now $x^{\mathrm{H}} = [\bar{x}_1 \;\; \bar{x}_2]$ can be complex

$$a|x_1|^2 + 2\mathrm{Re}\, b\bar{x}_1 x_2 + c|x_2|^2 = a|x_1 + (b/a)x_2|^2 + \underline{\qquad}|x_2|^2.$$

(c) Show that $a > 0$ and $ac > |b|^2$ ensure that A is positive definite.

(d) Are the matrices $\left[\begin{smallmatrix} 1 & 1+i \\ 1-i & 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 3 & 4+i \\ 4-i & 6 \end{smallmatrix}\right]$ positive definite?

12. Decide whether $F = x^2y^2 - 2x - 2y$ has a minimum at the point $x = y = 1$ (after showing that the first derivatives are zero at that point).

13. Under what conditions on a, b, c is $ax^2 + 2bxy + cy^2 > x^2 + y^2$ for all x, y?

Problems 14–18 are about tests for positive definiteness.

14. Which of A_1, A_2, A_3, A_4 has two positive eigenvalues? Test $a > 0$ and $ac > b^2$, don't compute the eigenvalues. Find an x so that $x^{\mathrm{T}}A_1x < 0$.

$$A_1 = \begin{bmatrix} 5 & 6 \\ 6 & 7 \end{bmatrix} \qquad A_2 = \begin{bmatrix} -1 & -2 \\ -2 & -5 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 1 & 10 \\ 10 & 100 \end{bmatrix} \qquad A_4 = \begin{bmatrix} 1 & 10 \\ 10 & 101 \end{bmatrix}.$$

15. What is the quadratic $f = ax^2 + 2bxy + cy^2$ for each of these matrices? Complete the square to write f as a sum of one or two squares $d_1(\qquad)^2 + d_2(\qquad)^2$.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}.$$

16. Show that $f(x,y) = x^2 + 4xy + 3y^2$ does not have a minimum at $(0,0)$ even though it has positive coefficients. Write f as a *difference* of squares and find a point (x,y) where f is negative.

17. (*Important*) If A has *independent columns*, then $A^{\mathrm{T}}A$ is square and symmetric and invertible (Section 4.2). ***Rewrite*** $x^{\mathrm{T}}A^{\mathrm{T}}Ax$ ***to show why it is positive except when*** $x = 0$. Then $A^{\mathrm{T}}A$ is positive definite.

18. Test to see if $A^{\mathrm{T}}A$ is positive definite in each case:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}, \qquad A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

19. Find the 3 by 3 matrix A and its pivots, rank, eigenvalues, and determinant:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 4(x_1 - x_2 + 2x_3)^2.$$

20. For $F_1(x,y) = \frac{1}{4}x^4 + x^2y + y^2$ and $F_2(x,y) = x^3 + xy - x$, find the second derivative matrices A_1 and A_2:

$$A = \begin{bmatrix} \partial^2 F/\partial x^2 & \partial^2 F/\partial x \partial y \\ \partial^2 F/\partial y \partial x & \partial^2 F/\partial y^2 \end{bmatrix}.$$

A_1 is positive definite, so F_1 is concave up (= convex). Find the minimum point of F_1 and the saddle point of F_2 (look where first derivatives are zero).

21. The graph of $z = x^2 + y^2$ is a bowl opening upward. *The graph* of $z = x^2 - y^2$ is *a saddle*. The graph of $z = -x^2 - y^2$ is a bowl opening downward. What is a test on $F(x,y)$ to have a saddle at $(0,0)$?

22. Which values of c give a bowl and which give a saddle point for the graph of $z = 4x^2 + 12xy + cy^2$? Describe this graph at the borderline value of c.

6.2 Tests for Positive Definiteness

Which symmetric matrices have the property that $x^{\mathrm{T}}Ax > 0$ for all nonzero vectors x? There are four or five different ways to answer this question, and we hope to find all of them. The previous section began with some hints about the signs of eigenvalues. but that gave place to the tests on a, b, c:

$$b = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad \text{is positive definite when} \quad a > 0 \quad \text{and} \quad ac - b^2 > 0.$$

From those conditions, ***both eigenvalues are positive***. Their product $\lambda_1\lambda_2$ is determinant $ac - b^2 > 0$, so the eigenvalues are either both positive or both negative. They must be positive because their sum is the trace $a + c > 0$.

Looking at a and $ac - b^2$, it is even possible to spot the appearance of the ***pivots***. They turned up when we decomposed $x^{\mathrm{T}}Ax$ into a sum of squares:

$$\textbf{Sum of squares} \qquad ax^2 + 2bxy + cy^2 = a\left(x + \frac{b}{a}y\right)^2 + \frac{ac - b^2}{a}y^2. \tag{1}$$

Those coefficients a and $(ac - b^2)/a$ are the pivots for a 2 by 2 matrix. For larger matrices the pivots still give a simple test for positive definiteness: $x^{\mathrm{T}}Ax$ stays positive when n independent squares are multiplied by ***positive pivots***.

One more preliminary remark. The two parts of this hook were linked by the chapter on determinants. Therefore we ask what part determinants play. ***It is not enough to require that the determinant of*** A ***is positive***. If $a = c = -1$ and $b = 0$. then $\det A = 1$ but $A = -I =$ negative definite. The determinant test is applied not only to A itself, giving $ac - b^2 > 0$, but also to the 1 by 1 submatrix a in the upper left-hand corner.

The natural generalization will involve all n of the *upper left submatrices* of A:

$$A_1 = \begin{bmatrix} a_{11} \end{bmatrix}, \qquad A_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \cdots, \qquad A_n = A.$$

Here is the main theorem on positive definiteness, and a reasonably detailed proof:

6B Each of the following tests is a necessary and sufficient condition for the real symmetric matrix A to be ***positive definite***:

(I) $x^{\mathrm{T}}kx > 0$ for all nonzero real vectors x.

(II) All the eigenvalues of A satisfy $\lambda_i > 0$.

(III) All the upper left submatrices A_k have positive determinants.

(IV) All the pivots (without row exchanges) satisfy $d_k > 0$.

Proof. Condition I defines a positive definite matrix. Our first step shows that each eigenvalue will be positive:

$$\text{If} \quad Ax = \lambda x, \quad \text{then} \quad x^{\mathrm{T}}Ax = x^{\mathrm{T}}\lambda x = \lambda \|x\|^2.$$

A positive definite matrix has positive eigenvalues, since $x^{\mathrm{T}}Ax > 0$.

Now we go in the other direction. If all $\lambda_i > 0$, we have to prove $x^{\mathrm{T}}Ax > 0$ for every vector x (not just the eigenvectors). Since symmetric matrices have a full set of orthonormal eigenvectors, any x is a combination $c_1x_1 + \cdots + c_nx_n$. Then

$$Ax = c_1Ax_1 + \cdots + c_nAx_n = c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n.$$

Because of the orthogonality $x_i^{\mathrm{T}}x_i = 0$, and the normalization $x_i^{\mathrm{T}}x_i = 1$,

$$\begin{aligned} x^{\mathrm{T}}Ax &= \left(c_1x_1^{\mathrm{T}} + \cdots + c_nx_n^{\mathrm{T}}\right)\left(c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n\right) \\ &= c_1^2\lambda_1 + \cdots + c_n^2\lambda_n. \end{aligned} \tag{2}$$

If every $\lambda_i > 0$, then equation (2) shows that $x^{\mathrm{T}}Ax > 0$. Thus condition II implies condition I.

If condition I holds, so does condition III: The determinant of A is the product of the eigenvalues. And if condition I holds, we already know that these eigenvalues are positive. But we also have to deal with every upper left submatrix A_k. The trick is to look at all nonzero vectors whose last $n-k$ components are zero:

$$x^{\mathrm{T}}Ax = \begin{bmatrix} x_k^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^{\mathrm{T}}A_kx_k > 0.$$

Thus A_k is positive definite. Its eigenvalues (not the same λ_1!) must be positive. Its determinant is their product, so all upper left determinants are positive.

If condition III holds, so does condition IV: According to Section 4.4, the kth pivot d_k is the ratio of $\det A_k$ to $\det A_{k-1}$. If the determinants are all positive, so are the pivots.

If condition IV holds, so does condition I: We are given positive pivots, and must deduce that $x^{\mathrm{T}}Ax > 0$. This is what we did in the 2 by 2 case, by completing the square. The pivots were the numbers outside the squares. To see how that happens for symmetric matrices of any size, we go back to *elimination on a symmetric matrix*: $A = LDL^{\mathrm{T}}$.

Example 1. Positive pivots 2, $\frac{3}{2}$, and $\frac{4}{3}$:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} = LDL^{\mathrm{T}}.$$

I want to split $x^{\mathrm{T}}Ax$ into $x^{\mathrm{T}}LDL^{\mathrm{T}}x$:

$$\text{If} \quad x = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \quad \text{then} \quad L^{\mathrm{T}}x = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} u - \frac{1}{2}v \\ v - \frac{2}{3}w \\ w \end{bmatrix}.$$

So $x^{\mathrm{T}}Ax$ is a sum of squares with the pivots 2, $\frac{3}{2}$, and $\frac{4}{3}$ as coefficients:

$$x^{\mathrm{T}}Ax = (L^{\mathrm{T}}x)^{\mathrm{T}}D(L^{\mathrm{T}}x) = \mathbf{2}\left(u - \frac{\mathbf{1}}{\mathbf{2}}v\right)^2 + \frac{\mathbf{3}}{\mathbf{2}}\left(v - \frac{\mathbf{2}}{\mathbf{3}}w\right)^2 + \frac{\mathbf{4}}{\mathbf{3}}(w)^2.$$

Those positive pivots in D multiply perfect squares to make $x^{\mathrm{T}}Ax$ positive. Thus condition IV implies condition I, and the proof is complete.

□

It is beautiful that elimination and completing the square are actually the same. Elimination removes x_1 from all later equations. Similarly, the first square accounts for all terms in $x^{\mathrm{T}}Ax$ involving x_1. The sum of squares has the pivots outside. *The multipliers ℓ_{ij} are inside*! You can see the numbers $-\frac{1}{2}$ and $-\frac{2}{3}$ inside the squares in the example.

Every diagonal entry a_{ii} must be positive. As we know from the examples, however, it is far from sufficient to look only at the diagonal entries.

The pivots d_i are not to be confused with the eigenvalues. For a typical positive definite matrix, they are two completely different sets of positive numbers, In our 3 by 3 example, probably the determinant test is the easiest:

$$\textbf{Determinant test} \qquad \det A_1 = 2, \quad \det A_2 = 3, \quad \det A_3 = \det A = 4.$$

The pivots are the ratios $d_1 = 2$, $d_2 = \frac{3}{2}$, $d_3 = \frac{4}{3}$. Ordinarily the eigenvalue test is the longest computation. For this A we know the λ's are all positive:

$$\textbf{Eigenvalue test} \qquad \lambda_1 = 2 - \sqrt{2}, \quad \lambda_2 = 2, \quad \lambda_3 = 2 + \sqrt{2}.$$

Even though it is the hardest to apply to a single matrix, eigenvalues can be the most useful test for theoretical purposes. ***Each test is enough by itself***.

Positive Definite Matrices and Least Squares

I hope you will allow one more test for positive definiteness. It is already close. We connected positive definite matrices to pivots (Chapter 1), determinants (Chapter 4), and eigenvalues (Chapter 5). Now we see them in the least-squares problems in Chapter 3, coming from the rectangular matrices of Chapter 2.

The rectangular matrix will be R and the least-squares problem will be $Rx = b$. It has m equations with $m \geq n$ (square systems are included). *The least-square choice* $\widehat{x}$ *is the solution of* $R^{\mathrm{T}}R\widehat{x} = R^{\mathrm{T}}b$. That matrix $AR^{\mathrm{T}}R$ is not only symmetric but positive definite, as we now show—provided that the n columns of R are linearly independent:

6C The symmetric matrix A is positive definite if and only if

(V) There is a matrix R with independent columns such that $A = R^{\mathrm{T}}R$.

The key is to recognize $x^{\mathrm{T}}Ax$ *as* $x^{\mathrm{T}}R^{\mathrm{T}}Rx = (Rx)^{\mathrm{T}}(Rx)$. This squared length $\|Rx\|^2$ is positive (unless $x = 0$), because R has independent columns. (If x is nonzero then Rx is nonzero.) Thus $x^{\mathrm{T}}R^{\mathrm{T}}Rx > 0$ and $R^{\mathrm{T}}R$ is positive definite.

It remains to find an R For which $A = R^{\mathrm{T}}R$. We have almost done this twice already:

$$\textbf{Elimination} \qquad A = LDL^{\mathrm{T}} = (L\sqrt{D})(\sqrt{D}L^{\mathrm{T}}). \quad \text{So take } R = \sqrt{D}L^{\mathrm{T}}.$$

This ***Cholesky decomposition*** has the pivots split evenly between L and L^{T}.

$$\textbf{Eigenvalues} \qquad A = Q\Lambda Q^{\mathrm{T}} = (Q\sqrt{\Lambda})(\sqrt{\Lambda}Q^{\mathrm{T}}). \quad \text{So take } R = \sqrt{\Lambda}Q^{\mathrm{T}}. \tag{3}$$

A third possibility is $R = Q\sqrt{\Lambda}Q^{\mathrm{T}}$, the ***symmetric positive definite square root*** of A. There are many other choices, square or rectangular, and we can see why. If you multiply any R by *a matrix Q with orthonormal columns*, then $(QR)^{\mathrm{T}}(QR) = R^{\mathrm{T}}Q^{\mathrm{T}}QR = R^{\mathrm{T}}IR = A$. Therefore QR is another choice.

Applications of positive definite matrices are developed in my earlier book *Introduction to Applied Mathematics* and also the new *Applied Mathematics and Scientific Computing* (see `www.wellesleycambridge.com`). We mention that $Ax = \lambda Mx$ arises constantly in engineering analysis. If A and M are positive definite, this generalized problem is parallel to the familiar $Ax = \lambda x$, and $\lambda > 0$. M is a ***mass matrix*** for the *finite element method* in Section 6.4.

Semidefinite Matrices

The tests for semidefiniteness will relax $x^{\mathrm{T}}Ax > 0$, $\lambda > 0$, $d > 0$, and det > 0, to allow zeros to appear. The main point is to see the analogies with the positive definite case.

6D Each of the following tests is a necessary and sufficient condition for a symmetric matrix A to be ***positive semidefinite***:

(I′) $x^{\mathrm{T}}Ax \geq 0$ for all vectors x (this defines positive semidefinite).

(II′) All the eigenvalues of A satisfy $\lambda_i \geq 0$.

(III′) No principal submatrices have negative determinants.

(IV′) No pivots are negative.

(V′) There is a matrix R, possibly with dependent columns, such that $A = R^{\mathrm{T}}R$.

The diagonalization $A = Q\Lambda Q^{\mathrm{T}}$ leads to $x^{\mathrm{T}}Ax = x^{\mathrm{T}}Q\Lambda Q^{\mathrm{T}}x = y^{\mathrm{T}}\Lambda y$. If A has rank r, there are r nonzero λ's and r perfect squares in $y^{\mathrm{T}}\Lambda y = \lambda_1 y_1^2 + \cdots + \lambda_r y_r^2$.

Note. The novelty is that condition III′ applies to all the principal submatrices, not only those in the upper left-hand corner. Otherwise, we could not distinguish between two matrices whose upper left determinants were all zero:

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \text{ is positive semidefinite, and } \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \text{ is negative semidefinite.}$$

A row exchange comes with the same column exchange to maintain symmetry.

Example 2.

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \quad \text{is positive } \textit{semi}\text{definite, by all five tests:}$$

(I′) $x^{\mathrm{T}}Ax = (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2 \geq 0$ (zero if $x_1 = x_2 = x_3$).

(II′) The eigenvalues are $\lambda_1 = 0$, $\lambda_2 = \lambda_3 = 3$ (a zero eigenvalue).

(III′) $\det A = 0$ and smaller determinants are positive.

(IV′) $A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{2} & \frac{3}{2} \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \mathbf{0} \end{bmatrix}$ (missing pivot).

(V′) $A = R^{\mathrm{T}}R$ with dependent columns in R:

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \quad (1,1,1) \text{ in the nullspace.}$$

Remark. The conditions for semidefiniteness could also be deduced from the origin conditions I-V for definiteness by the following trick: Add a small multiple of the identity giving a positive definite matrix $A + \varepsilon I$. Then let ε approach zero. Since the determinants and eigenvalues depend continuously on ε, they will be positive until the very last moment. At $\varepsilon = 0$ they must still be nonnegative.

My class often asks about *unsymmetric* positive definite matrices. I never use that term. One reasonable definition is that the symmetric part $\frac{1}{2}(A+A^{\mathrm{T}})$ should be positive definite. That guarantees that *the real parts of the eigenvalues are positive*. But it is not necessary: $A = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$ has $\lambda > 0$ but $\frac{1}{2}(A+A^{\mathrm{T}}) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is indefinite.

If $Ax = \lambda x$, then $x^{\mathrm{H}}Ax = \lambda x^{\mathrm{H}}x$ and $x^{\mathrm{H}}A^{\mathrm{H}}x = \overline{\lambda}x^{\mathrm{H}}x$.

Adding, $\frac{1}{2}x^{\mathrm{H}}(A+A^{\mathrm{H}})x = (\mathrm{Re}\lambda)x^{\mathrm{H}}x > 0$, so that $\mathrm{Re}\lambda > 0$.

Ellipsoids in n Dimensions

Throughout this book, geometry has helped the matrix algebra. A linear equation produced a plane. The system $Ax = b$ gives an intersection of planes. Least squares gives a perpendicular projection. The determinant is the volume of a box. Now, for a positive definite matrix and its $x^{\mathrm{T}}Ax$, we finally get a figure that is curved. It is an *ellipse* in two dimensions, and an *ellipsoid* in n dimensions.

The equation to consider is $x^{\mathrm{T}}Ax = 1$. If A is the identity matrix, this simplifies to $x_1^2 + x_2^2 + \cdots + x_n^2 = 1$. This is the equation of the "unit sphere" in $\mathbf{R}^n$. If $A = 4I$, the sphere gets smaller. The equation changes to $4x_1^2 + \cdots + 4x_n^2 = 1$. Instead of $(1,0,\ldots,0)$, it goes through $(\frac{1}{2},0,\ldots,0)$. The center is at the origin, because if x satisfies $x^{\mathrm{T}}Ax = 1$, so does the opposite vector $-x$. The important step is to go from the identity matrix to a *diagonal matrix*:

Ellipsoid $\qquad$ For $A = \begin{bmatrix} 4 & & \\ & 1 & \\ & & \frac{1}{9} \end{bmatrix}$, $\quad$ the equation is $x^{\mathrm{T}}Ax = 4x_1^2 + x_2^2 + \frac{1}{9}x_3^2 = 1$.

Since the entries are unequal (and positive!) the sphere changes to an ellipsoid.

One solution is $x = (\frac{1}{2},0,0)$ along the first axis. Another is $x = (0,1,0)$. The major axis has the farthest point $x = (0,0,3)$. It is like a football or a rugby ball, but not quite—those are closer to $x_1^2 + x_2^2 + \frac{1}{2}x_3^2 = 1$. The two equal coefficients make them circular in the x_1-x_2 plane, and much easier to throw!

Now comes the final step, to allow nonzeros away from the diagonal of A.

Example 3. $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ and $x^{\mathrm{T}}Ax = 5u^2 + 8uv + 5v^2 = 1$. That ellipse is centered at $u = v = 0$, but the axes are not so clear. The off-diagonal 4s leave the matrix positive definite, but they rotate the ellipse—its axes no longer line up with the coordinate axes (Figure 6.2). We will show that ***the axes of the ellipse point toward the eigenvector of*** A. Because $A = A^{\mathrm{T}}$, those eigenvectors and axes are orthogonal. The *major* axis of the ellipse corresponds to the *smallest* eigenvalue of A.

To locate the ellipse we compute $\lambda_1 = 1$ and $\lambda_2 = 9$. The unit eigenvectors are $(1,-1)/\sqrt{2}$ and $(1,1)/\sqrt{2}$. Those are at 45° angles with the u-v axes, and they are lined up with the axes of the ellipse. The way to see the ellipse properly is to *rewrite*

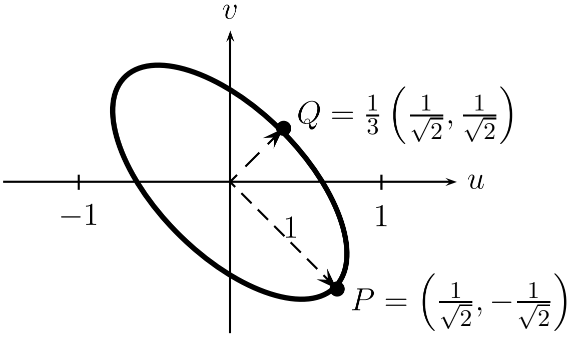

Figure 6.2: The ellipse $x^{\mathrm{T}}Ax = 5u^2 + 8uv + 5v^2 = 1$ and its principal axes.

$x^{\mathrm{T}}Ax = 1$:

$$\textbf{New squares} \qquad 5u^2 + 8uv + v^2 = \left(\frac{u}{\sqrt{2}} - \frac{v}{\sqrt{2}}\right)^2 + 9\left(\frac{u}{\sqrt{2}} + \frac{v}{\sqrt{2}}\right)^2 = 1. \tag{4}$$

$\lambda = 1$ and $\lambda = 9$ are outside the squares. The eigenvectors are inside. This is different from completing the square to $5(u + \frac{4}{5}v)^2 + \frac{9}{5}v^2$, with the *pivots* outside.

The first square equals 1 at $(1/\sqrt{2}, -1/\sqrt{2})$ at the end of the major axis. The minor axis is one-third as long, since we need $(\frac{1}{3})^2$ to cancel the 9.

Any ellipsoid $x^{\mathrm{T}}Ax = 1$ can be simplified in the same way. *The key step is to diagonalize $A = Q\Lambda Q^{\mathrm{T}}$.* We straightened the picture by rotating the axes. Algebraically, the change to $y = Q^{\mathrm{T}}x$ produces a sum of squares:

$$x^{\mathrm{T}}Ax = (x^{\mathrm{T}}Q)\Lambda(Q^{\mathrm{T}}x) = y^{\mathrm{T}}\Lambda y = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = 1. \tag{5}$$

The *major axis* has $y_1 = 1/\sqrt{\lambda_1}$ along the eigenvector with the smallest eigenvalue.

The other axes are along the other eigenvectors. Their lengths are $1/\sqrt{\lambda_2}, \ldots, 1/\sqrt{\lambda_n}$. Notice that the λ's must be positive—*the matrix must be positive definite*—or these square roots are in trouble. An indefinite equation $y_1^2 - 9y_2^2 = 1$ describes a hyperbola and not an ellipse. A hyperbola is a cross-section through a saddle, and an ellipse is a cross-section through a bowl.

The change from x to $y = Q^{\mathrm{T}}x$ rotates the axes of the space, to match the axes of the ellipsoid. In the y variables we can see that it is an ellipsoid, because the equation becomes so manageable:

> **6E** Suppose $A = Q\Lambda Q^{\mathrm{T}}$ with $\lambda_i > 0$. Rotating $y = Q^{\mathrm{T}}x$ simplifies $x^{\mathrm{T}}Ax = 1$:
>
> $$x^{\mathrm{T}}Q\Lambda Q^{\mathrm{T}}x = 1, \qquad y^{\mathrm{T}}\Lambda y = 1, \quad \text{and} \quad \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = 1.$$
>
> This is the equation of an ellipsoid. Its axes have lengths $1/\sqrt{\lambda_1}, \ldots, 1/\sqrt{\lambda_n}$ from the center. In the original x-space they point along the eigenvectors of A.

The Law of Inertia

For elimination and eigenvalues, matrices become simpler by elementary operations The essential thing is to know which properties of the matrix stay unchanged. When a multiple of one row is subtracted from another, the row space, nullspace. rant and determinant all remain the same. For eigenvalues, the basic operation was a similarity transformation $A \to S^{-1}AS$ (or $A \to M^{-1}AM$). The eigenvalues are unchanged (and also the Jordan form). Now we ask the same question for symmetric matrices: *What are the elementary operations and their invariants for $x^{\mathrm{T}}Ax$?*

The basic operation on a quadratic form is to change variables. A new vector y is related to x by some nonsingular matrix, $x = Cy$. The quadratic form becomes $y^{\mathrm{T}}C^{\mathrm{T}}ACy$. This shows the fundamental operation on A:

$$\textbf{Congruence transformation} \qquad A \to C^{\mathrm{T}}AC \quad \text{for some nonsingular } C. \tag{6}$$

The symmetry of A is preserved, since $C^{\mathrm{T}}AC$ remains symmetric. The real question is, What other properties are shared by A and $C^{\mathrm{T}}AC$? The answer is given by Sylvester's ***law of inertia***.

> **6F** $C^{\mathrm{T}}AC$ has the same number of positive eigenvalues, negative eigenvalues, and zero eigenvalues as A.

The *signs* of the eigenvalues (and not the eigenvalues themselves) are preserved by a congruence transformation. In the proof, we will suppose that A is nonsingular. Then $C^{\mathrm{T}}AC$ is also nonsingular, and there are no zero eigenvalues to worry about. (Otherwise we can work with the nonsingular $A + \varepsilon I$ and $A - \varepsilon I$, and at the end let $\varepsilon \to 0$.)

Proof. We want to borrow a trick from topology. Suppose C is linked to an orthogonal matrix Q by a continuous chain of nonsingular matrices $C(t)$. At $t = 0$ and $t = 1$, $C(0) = C$ and $C(1) = Q$. Then the eigenvalues of $C(t)^{\mathrm{T}}AC(t)$ will change gradually, as t goes from 0 to 1, from the eigenvalues of $C^{\mathrm{T}}AC$ to the eigenvalues of $Q^{\mathrm{T}}AQ$. Because $C(t)$ is never singular, *none of these eigenvalues can touch zero* (not to mention cross over it!). Therefore the number of eigenvalues to the right of zero, and the number to the left, is the same for $C^{\mathrm{T}}AC$ as for $Q^{\mathrm{T}}AQ$. And A has exactly the same eigenvalues as the similar matrix $Q^{-1}AQ = Q^{\mathrm{T}}AQ$.

One good choice for Q is to apply Gram-Schmidt to the columns of C. Then $C = QR$, and the chain of matrices is $C(t) = tQ + (1-t)QR$. The family $C(t)$ goes slowly through Gram-Schmidt, from QR to Q. It is invertible, because Q is invertible and the triangular factor $tI + (1-t)R$ has positive diagonal. That ends the proof. □

Example 4. Suppose $A = I$. Then $C^{\mathrm{T}}AC = C^{\mathrm{T}}C$ is positive definite. Both I and $C^{\mathrm{T}}C$ have n positive eigenvalues, confirming the law of inertia.

Example 5. If $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, then $C^{\mathrm{T}}AC$ has a negative determinant:

$$\det C^{\mathrm{T}}AC = (\det C^{\mathrm{T}})(\det A)(\det C) = -(\det C)^2 < 0.$$

Then $C^{\mathrm{T}}AC$ must have one positive and one negative eigenvalue, like A.

Example 6. This application is the important one:

> **6G** For any symmetric matrix A, ***the signs of the pivots agree with the signs of the eigenvalues***. The eigenvalue matrix Λ and the pivot matrix D have the same number of positive entries, negative entries, and zero entries.

We will assume that A allows the symmetric factorization $A = LDL^{\mathrm{T}}$ (without row exchanges). By the law of inertia, A has the same number of positive eigenvalues as D. But the eigenvalues of D are just its diagonal entries (the pivots). Thus the number of positive pivots matches the number of positive eigenvalues of A.

That is both beautiful and practical. It is beautiful because it brings together (for symmetric matrices) two parts of this book that were previously separate: *pivots* and *eigenvalues*. It is also practical, because the pivots can locate the eigenvalues:

$$\begin{array}{l}A \textbf{ has positive pivots}\\ A-2I \textbf{ has a negative pivot}\end{array} \qquad A = \begin{bmatrix}3 & 3 & 0\\ 3 & 10 & 7\\ 0 & 7 & 8\end{bmatrix} \qquad A-2I = \begin{bmatrix}1 & 3 & 0\\ 3 & 8 & 7\\ 0 & 7 & 6\end{bmatrix}.$$

A has positive eigenvalues, by our test. But we know that $\lambda_{\min}$ *is smaller than* 2, because subtracting 2 dropped it below zero. The next step looks at $A-I$, to see if $\lambda_{\min} < 1$. (It is, because $A-I$ has a negative pivot.) That interval containing λ is cut in half at every step by checking the signs of the pivots.

This was almost the first practical method of computing eigenvalues. It was dominant about 1960, after one important improvement—to make A tridiagonal first. Then the pivots are computed in $2n$ steps instead of $\frac{1}{6}n^3$. Elimination becomes fast, and the search for eigenvalues (by halving the intervals) becomes simple. The current favorite is the QR method in Chapter 7.

The Generalized Eigenvalue Problem

Physics, engineering, and statistics are usually kind enough to produce symmetric matrices in their eigenvalue problems. ***But sometimes*** $Ax = \lambda x$ ***is replaced by*** $Ax = \lambda Mx$. *There are two matrices rather than one.*

An example is the motion of two unequal masses in a line of springs:

$$\begin{aligned} m_1\frac{d^2v}{dt^2} + 2v - w &= 0\\ m_2\frac{d^2w}{dt^2} - v + 2w &= 0\end{aligned} \quad \text{or} \quad \begin{bmatrix}m_1 & 0\\ 0 & m_2\end{bmatrix}\frac{d^2u}{dt^2} + \begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}u = 0. \tag{7}$$

When the masses were equal, $m_1 = m_2 = 1$, this was the old system $u'' + Au = 0$. Now it is $Mu'' + Au = 0$, with a *mass matrix* M. The eigenvalue problem arises when we look for exponential solutions $e^{i\omega t}x$:

$$Mu'' + Au = 0 \qquad \text{becomes} \qquad M(i\omega)^2 e^{i\omega t}x + Ae^{i\omega t}x = 0. \tag{8}$$

Canceling $e^{i\omega t}$, and writing λ for ω^2, this is an eigenvalue problem:

$$\textbf{Generalized problem } Ax = \lambda Mx \qquad \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} x = \lambda \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} x. \tag{9}$$

There is a solution when $A - \lambda M$ is singular. The special choice $M = I$ brings back the usual $\det(A - \lambda I) = 0$. We work out $\det(A - \lambda M)$ with $m_1 = 1$ and $m_2 = 2$:

$$\det \begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-2\lambda \end{bmatrix} = 2\lambda^2 - 6\lambda + 3 = 0 \quad \text{gives} \quad \lambda = \frac{3 \pm \sqrt{3}}{2}.$$

For the eigenvector $x_1(\sqrt{3}-1, 1)$, the two masses oscillate together—but the first mass only moves as far as $\sqrt{3}-1 \approx .73$. In the fastest mode, the components of $x_2 = (1+\sqrt{3}, -1)$ have opposite signs and the masses move in opposite directions. This time the smaller mass goes much further.

The underlying theory is easier to explain if M is split into $R^{\mathrm{T}}R$. (*M is assumed to be positive definite.*) Then the substitution $y = Rx$ changes

$$Ax = \lambda Mx = \lambda R^{\mathrm{T}}Rx \quad \text{into} \quad AR^{-1}y = \lambda R^{\mathrm{T}}y.$$

Writing C for R^{-1}, and multiplying through by $(R^{\mathrm{T}})^{-1} = C^{\mathrm{T}}$, this becomes a standard eigenvalue problem for the *single* symmetric matrix $C^{\mathrm{T}}AC$:

$$\textbf{Equivalent problem} \qquad C^{\mathrm{T}}ACy = \lambda y. \tag{10}$$

The eigenvalues λ_j are the same as for the original $Ax = \lambda Mx$. and the eigenvectors are related by $y_j = Rx_j$. The properties of $C^{\mathrm{T}}AC$ lead directly to thc properties of $Ax = \lambda Mx$, when $A = A^{\mathrm{T}}$ and M is positive definite:

1. The eigenvalues for $Ax = \lambda Mx$ are real, because $C^{\mathrm{T}}AC$ is symmetric.

2. The λ's have the same signs as the eigenvalues of A, by the law of inertia.

3. $C^{\mathrm{T}}AC$ has orthogonal eigenvectors y_j. So the eigenvectors of $Ax = \lambda Mx$ have

$$\textbf{“}M\textbf{-orthogonality”} \qquad x_i^{\mathrm{T}}Mx_j = x_i^{\mathrm{T}}R^{\mathrm{T}}Rx_j = y_i^{\mathrm{T}}y_j = 0. \tag{11}$$

A and M are being *simultaneously diagonalized*. If S has the x_j in its columns, then $S^{\mathrm{T}}AS = \Lambda$ and $S^{\mathrm{T}}MS = I$. This is a *congruence* transformation, with S^{T} on the left, and not a similarity transformation with S^{-1}. The main point is easy to summarize: As long as M is positive definite, the generalized eigenvalue problem $Ax = -\lambda Mx$ behaves exactly like $Ax = \lambda x$.

Problem Set 6.2

1. For what range of numbers a and b are the matrices A and B positive definite?

$$A = \begin{bmatrix} a & 2 & 2 \\ 2 & a & 2 \\ 2 & 2 & a \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 4 \\ 2 & b & 8 \\ 4 & 8 & 7 \end{bmatrix}.$$

2. Decide for or against the positive definiteness of

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}^2.$$

3. Construct an indefinite matrix with its largest entries on the main diagonal:

$$A = \begin{bmatrix} 1 & b & -b \\ b & 1 & b \\ -b & b & 1 \end{bmatrix} \quad \text{with } |b| < 1 \text{ can have } \det A < 0.$$

4. Show from the eigenvalues that if A is positive definite, so is A^2 and so is A^{-1}.

5. *If A and B are positive definite, then $A+B$ is positive definite*. Pivots and eigenvalues are not convenient for $A+B$. Much better to prove $x^{\mathrm{T}}(A+B)x > 0$.

6. From the pivots, eigenvalues, and eigenvectors of $A = \left[\begin{smallmatrix} 5 & 4 \\ 4 & 5 \end{smallmatrix}\right]$, write A as $R^{\mathrm{T}}R$ in three ways: $(L\sqrt{D})(\sqrt{D}L^{\mathrm{T}})$, $(Q\sqrt{\Lambda})(\sqrt{\Lambda}Q^{\mathrm{T}})$, and $(Q\sqrt{\Lambda}Q^{\mathrm{T}})(Q\sqrt{\Lambda}Q^{\mathrm{T}})$.

7. If $A = Q\Lambda Q^{\mathrm{T}}$ is symmetric positive definite, then $R = Q\sqrt{\Lambda}Q^{\mathrm{T}}$ is its *symmetric positive definite square root*. Why does R have positive eigenvalues? Compute R and verify $R^2 = A$ for

$$A = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix}.$$

8. If A is symmetric positive definite and C is nonsingular, prove that $B = C^{\mathrm{T}}AC$ is also symmetric positive definite.

9. If $A = R^{\mathrm{T}}R$ prove the generalized Schwarz inequality $|x^{\mathrm{T}}Ay|^2 \le (x^{\mathrm{T}}Ax)(y^{\mathrm{T}}Ay)$.

10. The ellipse $u^2 + 4v^2 = 1$ corresponds to $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 4 \end{smallmatrix}\right]$. Write the eigenvalues and eigenvectors, and sketch the ellipse.

11. Reduce the equation $3u^2 - 2\sqrt{2}uv + 2v^2 = 1$ to a sum of squares by finding the eigenvalues of the corresponding A, and sketch the ellipse.

12. In three dimensions, $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 = 1$ represents an ellipsoid when all $\lambda_i > 0$. Describe all the different kinds of surfaces that appear in the positive semidefinite case when one or more of the eigenvalues is zero.

13. Write down the five conditions for a 3 by 3 matrix to be *negative definite* ($-A$ is positive definite) with special attention to condition III: How is $\det(-A)$ related to $\det A$?

14. Decide whether the following matrices are positive definite, negative definite, semidefinite, or indefinite:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 3 & 4 & 9 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 2 & 6 & -2 & 0 \\ 0 & -2 & 5 & -2 \\ 0 & 0 & -2 & 3 \end{bmatrix}, \qquad C = -B, \qquad D = A^{-1}.$$

Is there a real solution to $-x^2 - 5y^2 - 9z^2 - 4xy - 6xz - 8yz = 1$?

15. Suppose A is symmetric positive definite and Q is an orthogonal matrix. True or false:

(a) $Q^{\mathrm{T}}AQ$ is a diagonal matrix.

(b) $Q^{\mathrm{T}}AQ$ is symmetric positive definite.

(c) $Q^{\mathrm{T}}AQ$ has the same eigenvalues as A.

(d) e^{-A} is symmetric positive definite.

16. If A is positive definite and a_{11} is increased, prove from cofactors that the determinant is increased. Show by example that this can fail if A is indefinite.

17. From $A = R^{\mathrm{T}}R$. show for positive definite matrices that $\det A \le a_{11}a_{22}\cdots a_{nn}$. (The length squared of column j of R is a_{jj}. Use determinant = volume.)

18. (Lyapunov test for stability of M) Suppose $AM + M^{\mathrm{H}}A = -I$ with positive definite A. If $Mx = \lambda x$ show that $\mathrm{Re}A < 0$. (*Hint*: Multiply the first equation by x^{H} and x.)

19. Which 3 by 3 symmetric matrices A produce these functions $f = x^{\mathrm{T}}Ax$? Why is the first matrix positive definite but not the second one?

(a) $f = 2(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3)$.

(b) $f = 2(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_1x_3 - x_2x_3)$.

20. Compute the three upper left determinants to establish positive definiteness. Verify that their ratios give the second and third pivots.

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 5 & 3 \\ 0 & 3 & 8 \end{bmatrix}.$$

21. A positive definite matrix cannot have a zero (or even worse, a negative number) on its diagonal. Show that this matrix fails to have $x^{\mathrm{T}}Ax > 0$:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 4 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \text{is not positive when} \quad (x_1, x_2, x_3) = (\quad , \quad , \quad).$$

22. A diagonal entry a_{jj} of a symmetric matrix cannot be smaller than all λ's. If it were, then $A - a_{jj}I$ would have ____ eigenvalues and would be positive definite. But $A - a_{jj}I$ has a ____ on the main diagonal.

23. Give a quick reason why each of these statements is true:

(a) Every positive definite matrix is invertible.

(b) The only positive definite projection matrix is $P = I$.

(c) A diagonal matrix with positive diagonal entries is positive definite.

(d) A symmetric matrix with a positive determinant might not be positive definite!

24. For which s and t do A and B have all $\lambda > 0$ (and are therefore positive definite)?

$$A = \begin{bmatrix} s & -4 & -4 \\ -4 & s & -4 \\ -4 & -4 & s \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} t & 3 & 0 \\ 3 & t & 4 \\ 0 & 4 & t \end{bmatrix}.$$

25. You may have seen the equation for an ellipse as $(\frac{x}{a})^2 + (\frac{y}{b})^2 = 1$. What are a and b when the equation is written as $\lambda_1 x^2 + \lambda_2 y^2 = 1$? The ellipse $9x^2 + 16y^2 = 1$ has half-axes with lengths $a =$ ____, and $b =$ ____.

26. Draw the tilted ellipse $x^2 + xy + y^2 = 1$ and find the half-lengths of its axes from the eigenvalues of the corresponding A.

27. With positive pivots in D, the factorization $A = LDL^{\mathrm{T}}$ becomes $L\sqrt{D}\sqrt{D}L^{\mathrm{T}}$. (Square roots of the pivots give $D = \sqrt{D}\sqrt{D}$.) Then $C = L\sqrt{D}$ yields the ***Cholesky factorization*** $A = CC^{\mathrm{T}}$, which is "symmetrized LU":

$$\text{From} \quad C = \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix} \quad \text{find } A. \qquad \text{From} \quad A = \begin{bmatrix} 4 & 8 \\ 8 & 25 \end{bmatrix} \quad \text{find } C.$$

28. In the Cholesky factorization $A = CC^{\mathrm{T}}$, with $C = L\sqrt{D}$, the square roots of the pivots are on the diagonal of C. Find C (lower triangular) for

$$A = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 8 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 7 \end{bmatrix}.$$

29. The symmetric factorization $A = LDL^{\mathrm{T}}$ means that $x^{\mathrm{T}}Ax = x^{\mathrm{T}}LDL^{\mathrm{T}}x$:

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ b/a & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & (ac-b^2)/a \end{bmatrix} \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The left-hand side is $ax^2 + 2bxy + cy^2$. The right-hand side is $a(x + \frac{b}{a}y)^2 + \underline{\qquad} y^2$. The second pivot completes the square! Test with $a = 2$, $b = 4$, $c = 10$.

30. Without multiplying $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, find

(a) the determinant of A. (b) the eigenvalues of A.
(c) the eigenvectors of A. (d) a reason why A is symmetric positive definite.

31. For the semidefinite matrices

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \text{ (rank 2)} \qquad \text{and} \qquad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \text{ (rank 1)},$$

write $x^{\mathrm{T}}Ax$ as a sum of two squares and $x^{\mathrm{T}}Bx$ as one square.

32. Apply any three tests to each of the matrices

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 2 \end{bmatrix},$$

to decide whether they are positive definite, positive semidefinite, or indefinite.

33. For $C = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, confirm that $C^{\mathrm{T}}AC$ has eigenvalues of the same signs as A. Construct a chain of nonsingular matrices $C(t)$ linking C to an orthogonal Q. Why is it impossible to construct a nonsingular chain linking C to the identity matrix?

34. If the pivots of a matrix are all greater than 1, are the eigenvalues all greater than 1? Test on the tridiagonal -1, 2, -1 matrices.

35. Use the pivots of $A - \frac{1}{2}I$ to decide whether A has an eigenvalue smaller than $\frac{1}{2}$:

$$A - \frac{1}{2}I = \begin{bmatrix} 2.5 & 3 & 0 \\ 3 & 9.5 & 7 \\ 0 & 7 & 7.5 \end{bmatrix}.$$

36. An algebraic proof of the law of inertia starts with the orthonormal eigenvectors $x_1, \dots, x_p$ of A corresponding to eigenvalues $\lambda_i > 0$. and the orthonormal eigenvectors $y_1, \dots, y_q$ of $C^{\mathrm{T}}AC$ corresponding to eigenvalues $\mu_i < 0$.

(a) To prove that the $p+q$ vectors $x_1,\dots,x_p$, $Cy_1,\dots,Cy_q$ are independent, assume that some combination gives zero:

$$a_1x_1+\cdots+a_px_p = b_1Cy_1+\cdots+b_qCy_q \quad (=z,\text{ say}).$$

Show that $z^{\mathrm{T}}Az = \lambda_1 a_1^2+\cdots+\lambda_p a_p^2 \geq 0$ and $z^{\mathrm{T}}Az = \mu_1 b_1^2+\cdots+\mu_q b_q^2 \leq 0$.

(b) Deduce that the a's and b's are zero (proving linear independence). From that deduce $p+q \leq n$.

(c) The same argument for the $n-p$ negative λ's and the $n-q$ positive μ's gives $n-p+n-q \leq n$. (We again assume no zero eigenvalues—which are handled separately). Show that $p+q=n$, so the number p of positive λ's equals the number $n-q$ of positive μ's—which is the law of inertia.

37. If C is nonsingular, show that A and $C^{\mathrm{T}}AC$ have the same rank. Thus they have the same number of zero eigenvalues.

38. Find by experiment the number of positive, negative, and zero eigenvalues of

$$A = \begin{bmatrix} I & B \\ B^{\mathrm{T}} & 0 \end{bmatrix}$$

when the block B (of order $\frac{1}{2}n$) is nonsingular.

39. Do A and $C^{\mathrm{T}}AC$ always satisfy the law of inertia when C is not square?

40. In equation (9) with $m_1 = 1$ and $m_2 = 2$, verify that the normal modes are M-orthogonal: $x_1^{\mathrm{T}}Mx_2 = 0$.

41. Find the eigenvalues and eigenvectors of $Ax = \lambda Mx$:

$$\begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix} x = \frac{\lambda}{18}\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix} x.$$

42. If the symmetric matrices A and M are indefinite, $Ax = \lambda Mx$ might not have real eigenvalues. Construct a 2 by 2 example.

43. A *group* of nonsingular matrices includes AB and A^{-1} if it includes A and B. "Products and inverses stay in the group." Which of these sets are groups? *Positive definite symmetric matrices A, orthogonal matrices Q, all exponentials e^{tA} of a fixed matrix A, matrices P with positive eigenvalues, matrices D with determinant* 1. Invent a group containing only positive definite matrices.